GENE TARGETING IN THE MOUSE: INTRODUCING SPECIFIC MUTATIONS ASSOCIATED WITH CYSTIC FIBROSIS

WENDY LOUISE KIMBER

PhD
THE UNIVERSITY OF EDINBURGH
1994

MRC HUMAN GENETICS UNIT
EDINBURGH

I declare that this thesis has been composed by myself, and that all of the work is my own unless otherwise stated.

Wendy L Kimber

December 1994

ABSTRACT

Cystic fibrosis (CF) is the most common fatal autosomal recessive genetic disease with over 50 000 people affected worldwide. It is characterised by defective chloride ion transport and excessive mucus production which have a range of clinical consequences varying in severity between individuals. Over 400 disease-causing mutations have been described to date, with one mutation, the ΔF508 mutation, accounting for approximately 70% of all CF mutant chromosomes. There appears to be little correlation between disease severity and genotype beyond pancreatic function, and the influence of environmental factors and genetic background is still to be determined.

Many experiments have been conducted both *in vitro* and *ex vivo*, investigating the pathophysiology of this disease. However, CF is a complex systemic disease affecting different organs and tissues to varying degrees, and an animal model is therefore of great value. Various mouse models have been described for CF, in which most or all *Cftr* gene expression is abolished. Whilst these animals are valuable to CF research, they do not provide answers to fundamental questions pertaining to the phenotype/genotype relationship, and the mechanisms by which different mutations in *Cftr* culminate in the wide spectrum of clinical features. In addition, animal models bearing mutations which give rise to mislocalised but partially functional mutant protein, including the ΔF508 mutation, would be invaluable for the testing and development of relocation strategies.

This thesis describes an attempt to create mouse models for CF bearing precise, CF-associated mutations as the only alteration to the murine *Cftr* gene, through use of the 'Hit and Run' gene targeting technique in mouse embryonal stem (ES) cells. This involves an initial targeting step in which an insertional targeting vector bearing the mutation to be introduced integrates through homologous recombination into *Cftr*. A negative selection then follows, enriching for those cells which have subsequently excised the vector and either reverted to the wild type genotype, or have been

converted to the desired genotype of the introduced mutation in the correct location as the only alteration to the gene.

Vectors incorporating CF-associated mutations were introduced into ES cells and found to target *Cftr* at a higher frequency than had been previously reported for this region of the murine CF gene. Selection against the integrated vector in these targeted 'hit' clones was conducted with the unexpected result of 100% of resistant clones retaining the vector and the selection cassette. Close examination of these clones discovered no gross rearrangements, but a failure of methylation-sensitive restriction enzymes to cut within this region which could be induced or removed by growing the clone in negative or positive selection respectively. Further investigation of this phenomenon led to the conclusion that expression of the selection cassette was being modulated by methylation, and loss of the negative selection gene expression by this mechanism was occurring far more frequently than the desired loss through vector excision. The 'run' procedure was modified in view of this to reduce the high methylation-induced background, and clones were subsequently obtained which had survived the selection through vector excision.

The methylation hypothesis provides an explanation for the unsuccessful attempts in this lab and perhaps others, to introduce subtle mutations into *Cftr* utilising techniques which involve a negative selection step. A modification of the 'hit and run' procedure which takes account of this phenomenon has demonstrated that it should be possible to overcome the problems encountered to date, and successfully generate mouse models for CF bearing clinically relevant mutations.

ABBREVIATIONS

A	Adenine
ABC	ATP binding cassette
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
APRT	Adenosine phosphoribosyltransferase enzyme
aprt	Adenosine phosphoribosyltransferase gene
ATCC	American type culture collection
ATP	Adenosine triphosphate
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
cAMP	Cyclic adenosine monophosphate
CBAVD	Congenital bilateral absence of the vas deferens
cDNA	Complementary DNA
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator protein
CFTR	Human cystic fibrosis transmembrane conductance regulator gene
Cftr	Murine cystic fibrosis transmembrane conductance regulator gene
CHO cells	Chinese hamster ovary cells
CONC	Concentration
CpG	Cytosine guanine dinucleotide
d.p.c.	Days post coitum
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DIA	Differentiation inhibitory activity
DMEM	Dulbecco's modification of Eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dTTP	Deoxythymidine triphosphate
ER	Endoplasmic reticulum
ES cells	Embryonic stem cells
FCS	Foetal calf serum
FIAU	1(1-2-deoxy-2-fluoro-β-D-arabinofuransyl)-5-iodouracil
FREQ.	Frequency
G	Guanine
G	Gancyclovir
G418	Geneticin, also known as Neomycin sulphate
GANC	Gancyclovir
$GANC^R$	Gancyclovir resistant
GMEM	Glasgow's modification of Eagle's medium
H	*Hpa II*
H	Hygromycin

HCG	Human chorionic gonadotrophin
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HPRT	Hypoxanthine phosphoribosyl transferase
hsvtk	Herpes simplex virus thymidine kinase gene
HYG	Hygromycin
Hyg/*tk*	Hygromycin-Thymidine kinase
HYG^R	Hygromycin resistant
i.p.	Intra-peritoneal
i.u.	International unit
IRT	Immunoreactive trypsin
kb	Kilobase
kd	Kilodalton
LIF	Leukaemia inhibitory factor
M	*Msp I*
MDR	Multiple drug resistant
MeCPs	Methyl CpG-binding protein
MEM	Modification of Eagle's medium
mRNA	Messenger RNA
MSD	Membrane spanning domain
NBD	Nucleotide binding domain
NC	Not counted
Neo	Neomycin
OD	Optical density
ORCC	Outwardly-rectifying chloride channel
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Potential difference
PEF	Primary embryonic fibroblasts
PI	Pancreatic insufficient
PMSG	Pregnant mare serum gonadotrophin
PNK	Polynucleotide kinase
PNS	Positive-negative selection
PS	Pancreatic sufficient
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
secs	Seconds
T	Thymidine
Temp	Temperature
tk	Thymidine kinase gene
T_M	Melting temperature
TMTC	Too many to count
u.v.	Ultra violet
wt	Wildtype
X-S	*Xba I-Sal I* restriction fragment
X-X	*Xba I* restriction fragment

Mutation nomenclature:

Mutation type:

Δ	deletion
ins	insertion
→	splice site alteration
X	stop

Information order:

first letter(s)	original amino acid
number	position of affected amino acid
last letter(s)	amino acid alteration

Examples:	ΔF508	deletion of phenylalanine at position 508
	R117H	substitution of arginine with histidine at 117
	435insA	insertion alanine at 435
	621+1G→T	G to T base substitution at 621+1

ACKNOWLEDGEMENTS

I wish to thank David Porteous and Julia Dorin for giving me the opportunity to study for a PhD on such an exciting project, and for the excellent training that I have received. I am grateful to David for all the time, guidance and support he has given me, and to Julia for being such a supportive, considerate, and conscientious supervisor, but also a good friend. I wish to express my gratitude to the Cystic Fibrosis Research Trust for their generous funding of this project.

Many people at the MRC Human Genetics unit have contributed to making my four and half years here a time of many fond memories. In particular I would like to thank the West Wingers past and present, for creating such a friendly and zany environment which I loved being part of. So many people deserve a special mention. Sheila Christie for not only running such a well organised lab, but for always coming to my rescue with sympathy, help and advice. Fiona Kilanowski for keeping me sane during many long, long hours of tissue culture, and for generally sharing the ups and downs of life. Many computer disasters have been avoided thanks to Chris 'Turbo' Boyd, and many lunchtimes filled with 'scientific' discussion. Kathy and Viv for all their advice and encouragement, and the benefit of following in their footsteps. Julia B for many laughs and culinary discussions, and Barbie for debate, intrigue, and good fun. Paul for unselfishly turning over his targeting vectors to me, and always being willing to help decipher a confusing Southern or a complicated digest. Thanks also to Sheila Webb for her patience when teaching me transgenic techniques.

Literature reviews would have been much more difficult if not for the extremely efficient library service provided by Sheila Mould. I am grateful to the Sandy, Douglas, and Norman for the great job they have made of all of my figures, and for the many slides and posters they have created for me in the past. I am also indebted to Ian Jackson for giving me my first experience of 'proper' research, and for the help and advice he has continued to offer even after I deserted him.

Heartfelt thanks also go to my family, especially mum, dad, and Nan, for the support and encouragement they have continually given. Their unshakeable belief in me has kept me going many times when things got tough. Special thanks must also go to my 'family-in-law', particularly Joy and Ken for their interest, encouragement, and support of my career which has meant so much.

Finally, I would like to dedicate this thesis to my husband Chris, for his patience, encouragement, help, and most of all, for the sacrifices he has made for me and my career. His love, and constant support has made this PhD a far less arduous task than it would otherwise have been.

CONTENTS

Declaration ii
Abstract iii
Abbreviations v
Acknowledgements viii
Contents x
List of Figures xv
List of Tables xiii

1. INTRODUCTION 1
1.1 Cystic Fibrosis 2
1.1.1 Clinical and Pathological Features 2
1.1.2 The Basic Defect In CF 5
1.1.3 Cloning The CF Gene 7
1.1.4 The Structure Of CFTR 8
1.1.5 The Function Of CFTR 9
1.1.6 CF Mutations 12
1.1.7 The Expression Of CFTR 20
1.1.8 Cross Species Analysis 21

1.2 Modelling Human Genetic Disease In An Animal System 22
1.2.1 The Value Of Animal Models 22
1.2.2 Choice Of Animal And Technology 23
1.2.3 Embryonal Stem Cell Technology 26
1.2.4 Gene Targeting 31
1.2.5 Animal Models For Human Genetic Disease Produced By Gene Targeting 43

1.3 Project Aim 49

2. MAMMALIAN CELL CULTURE 51
2.1 Embryonal Stem Cell Culture 52
2.1.2 Culture Media And Solutions 52
2.1.3 Production O f Leukaemia Inhibitory Factor (LIF) 54
2.1.4 Preparation Of Feeder Layers 55
2.1.5 Maintenance Of ES Cells 57
2.1.6 Karyotype Analysis Of ES Cells 59
2.1.7 Mycoplasma Detection 60
2.1.8 Electroporation Of ES Cells 62
2.1.9 Drug Selection 63
2.1.10 Clone Picking And Expansion 65
2.1.11 Long Term Storage Of ES Cells 66

2.2 Non-ES Cell Culture 67
2.2.1 Cell Line 67
2.2.2 Culture Media And Solutions 68

2.2.3 Maintenance Of Cos-7 Cells 68
2.2.4 Freezing And Long-term Storage Of Cos-7 Cells 68

3. MANIPULATION OF NUCLEIC ACIDS 69
3.1 Isolation Of Nucleic Acids From Mammalian Cells 70
3.1.1 Isolation Of DNA From Mammalian Cells 70
3.1.2 Quantification Of DNA 72

3.2 Bacterial Cell Culture 72
3.2.1 Bacterial Host Strains 72
3.2.2 Media And Additives 72
3.2.3 Preparation Of Competent Cells For Heat Shock Transformation 73
3.2.4 Heat Shock Transformation Of Bacterial Cells 74
3.2.5 Small Scale Plasmid Preparation 74
3.2.6 Large Scale Plasmid DNA Preparation 75

3.3 Analysis Of DNA By Restriction Enzyme Digestion 77
3.3.1 Plasmid DNA Digest 77
3.3.2 Genomic DNA Digest 77

3.4 Gel Electrophoresis Of Nucleic Acids 77
3.4.1 Solutions And Buffers 78
3.4.2 Agarose Gel Electrophoresis Of DNA 78
3.4.3 Preparative DNA Electrophoresis 79

3.5 Transfer Of Nucleic Acid To Membranes 80
3.5.1 Southern Transfer Of DNA. 80

3.6 Radiolabelling Of DNA 81
3.6.1 Random Primed Labelling Of DNA 81
3.6.2 Pre-annealing Of Repetitive Sequences 82
3.6.3 End Labelling Of DNA Oligonucleotides 82

3.7 Nucleic Acid Hybridisation 83
3.7.1 Hybridisation Solutions 83
3.7.2 Prehybridisation And Hybridisation Protocols 83
3.7.3 Post-hybridisation Washing Protocols 84
3.7.4 Removal Of Hybridisation Signal. 84

3.8 Detection Of Hybridisation 85
3.8.1 Autoradiography 85
3.8.2 Phosphorimaging 85

3.9 Amplification Of DNA By The Polymerase Chain Reaction (PCR). 86
3.9.1 PCR Of Genomic DNA 86

4. TRANSGENIC TECHNIQUES **89**
4.1 Technical Considerations 90
4.2 Preparation Of ES Cells For Chimaera Production 90
4.3 Mouse Strains 91
4.4 Superovulation Of Females For Embryo Collection 91
4.5 Preparation Of Pseudopregnant Foster Mothers 92
4.6 Medium And Solutions 92
4.7 Morula Collection 94
4.8 Aggregation Formation 94
4.9 Transfer Of Manipulated Embryos To Pseudopregnant Host Females 95
4.10 Analysis Of Progeny From Embryo Transfer 97

5. HIT AND RUN TARGETING WITH ONE STEP ANALYSIS **98**
5.1 Introduction **99**
5.1.1 Strategy 99

5.2 Hit And Run Targeting Vectors **102**
5.2.1 Vector Construction 102

5.3 Hit And Run Targeting **105**
5.3.1 Experiment HR1 105
5.3.2 Experiment HR2 111

5.4 Discussion **116**

6. TARGETING THE *Cftr* LOCUS **117**
6.1 Introduction **118**

6.2 HR3 Targeting Experiment **118**
6.2.1 Method 118
6.2.2 Results 118
6.2.3 Conclusions 126

6.3 HR5 Targeting Experiment **127**
6.3.1 Method 127
6.3.2 Results 127
6.3.3 Conclusions 133
6.3.4 Generating Transgenic Mice From Targeted 'Hit' Clones 133

6.4 Discussion **135**
6.4.1 Cell Lines 136
6.4.2 Targeting Frequencies 137
6.4.3 Structure Of Insertion Events 144
6.4.4 Loss Of Mutations During Integration 145

7. SELECTION FOR THE 'RUN' EVENT **149**
7.1 Introduction **150**

7.2 RUN2 'Run' Experiment **150**
7.2.1 Method 150
7.2.2 Results 151
7.2.3 Conclusions 158

7.3 RUN3 'Run' Experiment **162**
7.3.1 Method 162
7.3.2 Results 163
7.3.3 Conclusions 166

7.4 RUN4 'Run' Experiment **167**
7.4.1 Method 168
7.4.2 Results 168
7.4.3 Conclusions 171

7.5 RUN6 'Run' Experiment **174**
7.5.1 Method 174
7.5.2 Results 174
7.5.3 Conclusions 177

7.6 Discussion 179

8. INVESTIGATION OF THE GANCYCLOVIR RESISTANCE OF 'RUN' CLONES **182**
8.1 Introduction **183**

8.2 Investigation Of The Selection Resistant Phenotype Of RUN4 Clones **183**
8.2.1 Method 183
8.2.2 Results 184
8.2.3 Conclusions 186

8.3 Further Southern Blot Analysis Of RUN4 Clones **188**
8.3.1 Method 188
8.3.2 Results 189
8.3.3 Conclusions 192

8.4 Correlation Of *Sal I* Digestion With Fusion Gene Expression **193**
8.4.1 Method 194
8.4.2 Results 194
8.4.3 Conclusions 196

8.5 An Investigation Into The Methylation Status Of The RUN4 Clones **196**
8.5.1 Method 197
8.5.2 Results 197
8.5.3 Conclusions 203

8.6 Discussion **206**

9. MODIFICATION OF THE 'HIT AND RUN' TARGETING **211**
9.1 Introduction **212**

9.2 RUN7 'Run' Experiment **212**
9.2.1 Method 213
9.2.2 Results 215
9.2.3 Conclusions 220

9.3 Targeting *Cftr* With A Modified 'Hit and Run' Targeting Vector 225
9.3.1 Method 225
9.3.2 Results 227
9.3.3 Conclusions 233

9.4 Discussion **233**

10. CONCLUDING REMARKS **236**
10.1 Thesis Summary **237**
10.2 Discussion **239**
10.3 Overall Conclusions **241**
10.4 Future Directions **241**

REFERENCES **244**

APPENDIX 1 **271**

PUBLICATIONS **298**

LIST OF FIGURES

Chapter 1

Figure 1.1	The proposed structure of CFTR.	9
Figure 1.2	The biosynthesis and function of CFTR in an epithelial cell.	19
Figure 1.3	Generating transgenic mice from ES cells.	30
Figure 1.4	The double-strand-break repair model.	32
Figure 1.5	The integration patterns of insertion and replacement vectors.	35
Figure 1.6	The positive-negative selection (PNS) procedure.	37
Figure 1.7	The 'hit and run' gene targeting strategy.	41

Chapter 5

Figure 5.1	The 'hit and run' gene targeting strategy.	101
Figure 5.2	*Cftr* targeting vectors.	104
Figure 5.3	The possible genotypes of clones from 'hit and run' targeting.	107
Figure 5.4	Hybridisation of ΔF508 allele-specific oligonucleotide (M10ΔF) to PCR-amplified *Cftr* exon 10 DNA of HR1 clones.	109
Figure 5.5	Amplification of vector sequences in HR1 clone DNA.	110

Chapter 6

Figure 6.1	Restriction map of exon 10 target site.	118
Figure 6.2	Screening for targeted 'hit' clones.	119
Figure 6.3	Southern blot of three HR3 targeted 'hit' clones.	120
Figure 6.4	Confirmation of HR3 'hit' clones.	121
Figure 6.5	Screening HR3 'hit' clones for the presence of the ΔI506/7 mutation.	123
Figure 6.6	Southern blot of HR5 targeted 'hit' clones.	127
Figure 6.7	Confirmation of targeted clones by reprobing with a 5' probe.	128
Figure 6.8	Screening HR5 'hit' clones for the presence of the ΔI506/7 mutation.	129

Chapter 7

Figure 7.1	The structure and restriction map of targeted and wildtype alleles.	154
Figure 7.2	Southern blot analysis of RUN2 Gancyclovir resistant clones.	155
Figure 7.3	Screening RUN3 clones for the presence of the 'hit and run' targeting vector.	163
Figure 7.4	Screening RUN4 Clones for excision of the 'hit and run' targeting vector.	168
Figure 7.5	The number of Gancyclovir resistant clones obtained at two different plating densities in 'run' experiments RUN3 and RUN4.	171
Figure 7.6	Screening RUN6 clones for the presence of the 'hit and run' targeting vector.	174
Figure 7.7	A comparison of the number of Gancyclovir resistant clones obtained from two different 'hit' clones under identical conditions.	176

Chapter 8

Figure 8.1	Further Southern blot analysis of RUN4 clones.	188
Figure 8.2	Southern blot analysis of RUN4 clones grown on Hygromycin or Gancylovir selection	194
Figure 8.3	Analysis of the methylation status of the Hygromycin-*tk* fusion gene of RUN4 clones by Southern blot utilising methylation-sensitive restriction enzymes.	199
Figure 8.4	Analysis of the methylation status of the 3' sequences flanking the target site by Southern blot utilising methylation-sensitive restriction enzymes.	201
Figure 8.5	Analysis of the methylation status of the 5' sequences flanking the integrated targeting vector of RUN4 clones by Southern blot utilising methylation-sensitive restriction enzymes.	202
Figure 8.6	Southern blot analysis of the 3' region of the *Cftr* target site in different mouse strains using methylation-sensitive enzymes.	204

Chapter 9

Figure 9.1	Flow chart illustration of the modified 'run' selection procedure.	214
Figure 9.2	Screening RUN7 clones by Southern blot for vector excision.	217
Figure 9.3	Screening RUN7 clones by Southern blot analysis for retention of the ΔI506/7 mutation.	219

Figure 9.4	Duplicated regions available for intrachromosomal recombination during a 'run' event.	224
Figure 9.5	Distribution of genomic sequences of 'hit and run' targeting vectors.	226
Figure 9.6	Restriction map of *Cftr* alleles of clones targeted with pHRΔI.	228
Figure 9.7	Screening experiment HR6 for targeted 'hit' clones.	230
Figure 9.8	Screening HR6 'hit' clones for the presence of the ΔI506/7 mutation.	232

LIST OF TABLES

Chapter 1

Table 1.1	Classes of *CFTR* mutations that cause CF.	18

Chapter 5

Table 5.1	Numbers of clones obtained at each stage of experiment HR1.	106
Table 5.2	Numbers of clones obtained at each stage of experiment HR2.	113

Chapter 6

Table 6.1	Numbers of Hygromycin resistant clones obtained in HR3 targeting experiment.	117
Table 6.2	Numbers of Hygromycin resistant clones obtained in HR5 targeting experiment.	126
Table 6.3	Presence of ΔI506/7 mutation, and chromosome counts of HR5 'hit' clones.	130
Table 6.4	Targeting efficiency at the *Cftr* locus using insertional vectors.	136
Table 6.5	A comparison of reported targeting frequencies and vectors used to target similar regions around exon 10 of *Cftr*.	137

Chapter 7

Table 7.1	The conditions used for seeding cells for RUN2.	149
Table 7.2	The numbers of clones arising in experiment RUN2.	151
Table 7.3	A summary of the number of Gancyclovir resistant clones arising in experiment RUN2.	152
Table 7.4	Screening HR5 RUN3 clones for vector excision.	162
Table 7.5	The numbers of Gancyclovir resistant clones obtained in experiment RUN4.	167
Table 7.6	The number of Gancyclovir resistant clones obtained from experiment RUN6.	173

Chapter 8

Table 8.1	The selection resistant phenotype of RUN4 clones.	183
Table 8.2	The selection resistant phenotype of RUN4 clones and their correlation with the degree of cutting at the *Sal I* site of the 'hit and run' targeting vector.	189

Chapter 9

Table 9.1	The number of Gancyclovir resistant clones picked in experiment RUN7.	216

CHAPTER 1
INTRODUCTION

1.1 CYSTIC FIBROSIS

This thesis describes an attempt to create mouse models for cystic fibrosis (CF) by gene targeting. In contrast to previous CF mouse models, the mice generated in this study should bear precise mutations associated with the disease in man. The study of mice bearing subtle, clinically relevant CF mutations should allow further elucidation of many aspects of this multifactorial disease. The following introduction seeks to give an overview of the nature of CF, and provide some insight into the many aspects of the disease which remain unresolved. A more complete description is presented in appendix 1.

Cystic fibrosis (CF) is the most common fatal autosomal recessive genetic disease of the Caucasian population with about 1 in 2000 live births affected, and therefore a carrier frequency of around 1 in 22. It is characterised by defective chloride ion transport and excessive mucus production in exocrine tissues which leads to a variety of clinical consequences (Boat *et al.* 1989).

1.1.1 Clinical and Pathological Features

The range and severity of symptoms associated with CF varies considerably and consequently so does the age of diagnosis. Previously, CF patients did not usually survive their teenage years and were cared for by paediatric specialists. Improvements in the care and management of CF have led to a concurrent improvement in life expectancy and it has been projected that a CF patient born in 1990 can expect an average life span of 40 years, double that of 20 years ago (Elborn *et al.* 1991).

Meconium ileus, a failure to pass meconium following birth, occurs in 5-10% of CF cases and is virtually diagnostic for the disease. It is thought to result from failure of pancreatic secretion leading to non-digestion of intraluminal contents also perhaps combined with dehydration, resulting in a bowel obstruction. Beyond the newborn period, bowel obstruction with incompletely digested contents can also occur and is termed 'meconium ileus equivalent'.

In many cases of CF, an absence of pancreatic secretions is observed as a consequence of obstruction of the pancreatic ducts with inspissated mucus followed by dilation and ultimately destruction of acinar cells and their replacement with fibrous tissue. The lack of pancreatic digestive enzymes leads to a variety of digestive disorders such as steatorrhoea, which are often manifested as a failure to gain weight. Diabetes, although endocrinic in nature, is sometimes a complication of CF and has been attributed to disturbance of islet architecture by acinar destruction and fibrosis. Pancreatic insufficiency in CF is rarely life-threatening and can be managed to a certain extent by pancreatic enzyme supplementation.

Lung infection and the subsequent tissue damage are the major cause of premature death in CF (Elborn and Shale 1990). There is great variation in age of onset, severity and course of lung disease between affected individuals. Although antibiotic treatment may suppress infection to a certain extent, it cannot easily be eradicated and there is an inevitable decline in lung function with time. Lung disease in CF arises as a result of the production of excessive viscous mucus which cannot be cleared from the lungs and so provides an environment which is easily colonised by opportunistic pathogens. Histological abnormalities such as submucosal gland hypertrophy and mucus cell hyperplasia are detectable within the first few days of life, prior to micro-organism colonisation, indicating that mucus accumulation is the primary event. Repeated endobronchial infection initiates a continuous inflammatory response involving both humoral and cellular responses. Prolonged stimulation of host defences by micro-organisms and their antigenic by-products leads to the accumulation of cytotoxic by-products such as oxidants, proteases and lipid products, all of which are damaging to the pulmonary epithelia. Infection of this damaged tissue re-stimulates the inflammatory response leading to a vicious circle of infection, inflammation, and lung injury. Chronic pulmonary infection and lung injury eventually culminate in respiratory failure associated with pulmonary hypertension and cor pulmonale.

Fertility is much reduced in CF patients, especially in the male. Greater than 95% of males have altered Wolffian duct structures, commonly congenital bilateral absence of the vas deferens (CBAVD), and are therefore azoospermic (Denning *et al.* 1968., Kaplan *et al.* 1968). Female CF patients are generally more fertile than males with around 75% able to sustain pregnancies (Cohen *et al.* 1980). Poor health status and thick cervical mucus often hinder conception, and the physical stress imposed by pregnancy is often detrimental to the health of the mother (Boat 1989., Kotloff *et al.* 1992).

Other complications of CF can include cirrhosis of the liver and arthropathy (a form of joint disease). The psychosocial consequences of living with a life-threatening and incurable disease which imposes a demanding daily regime of physiotherapy and drug therapy are far reaching. The poor prognosis of patients contracting the highly infective *Pseudomonas cepacia* (recently renamed *Burkholdaria cepacia*), has led to the segregation of patients with positive sputum cultures both inside and outside of the hospital environment (Smith *et al.* 1993., Govan *et al.* 1993). This is particularly stressful, since patients had previously been encouraged to socialise, and many facilities and associations have been set up to encourage this. Living with infertility can also have psychological complications, especially if the patient is in relatively good health.

Diagnosis of CF is usually suggested from the clinical features and then confirmed by the sweat test (Gibson and Cooke 1959), and/or a test for immunoreactive trypsin (IRT) (Ryley *et al.* 1981). The concentration of chloride in sweat is determined with the sweat test, an elevated concentration being indicative (but not diagnostic) of CF. The IRT test was developed primarily for screening neonates where difficulty in obtaining sufficient quantities of sweat is often experienced. This uses immunoassay techniques to measure the presence of immunoreactive trypsin in the blood which is caused by leak-back of trypsin from an obstructed pancreas.

1.1.2 The Basic Defect In CF.

The first insight into the basic biochemical defect of CF arose in 1981 when Knowles *et al.* reported that the potential differences (PD) across nasal and airway epithelia in CF patients of all ages and stages of the disease were elevated, approximately twice that of normal controls. The application of amiloride, an inhibitor of active sodium absorption, resulted in a decrease in PD in both normal and CF epithelia. However this decrease was much greater in CF epithelia, indicating that sodium absorption was also elevated in CF. The role of chloride transport in CF became apparent in 1983 when Quinton demonstrated abnormally low chloride permeability in CF sweat ducts. This would result in poor sodium chloride reabsorption from sweat and thus accounts for the high sweat salt concentration characteristic of CF. This result was quickly followed by demonstration of reduced chloride permeability in CF nasal mucosa (Knowles *et al.* 1983).

Since these early studies, reduced chloride permeability has been shown in most affected epithelia (Welsh 1990), and increased sodium transport in airway epithelia (Knowles *et al.* 1986., Cotton *et al.* 1987). The defective chloride permeability was localised to the apical membrane (Widdicombe *et al.* 1985), and results *in vivo* confirmed by experimentation with primary and immortalised cultures of CF epithelial cells *in vitro* (Stutts *et al.* 1985., Jetten *et al.* 1989). Use of the patch-clamp technique to measure chloride conductance across epithelial membranes demonstrated that β-adrenergic-stimulated cAMP-induced activation of chloride channel activity was defective in CF (Frizzell *et al.* 1986). Further studies suggested that the block in cAMP-mediated activation of CF chloride channels lay distal to induction of cAMP-dependent protein kinase activation, and might be a defect in the chloride channel itself or an associated regulatory protein (Schoumacher *et al.* 1987., Li *et al.* 1988).

In September 1989, the study of CF changed dramatically with the cloning of the CF gene (Rommens *et al.* 1989., Riordan *et al.* 1989., Kerem, *et al.* 1989). Initially, homology to a family of transport proteins suggested that the gene product, the cystic fibrosis conductance regulator (CFTR), was itself a transporter and not a chloride

channel (Hyde *et al.* 1990., Ringe and Petsko, 1990). The subsequent availability of *CFTR* cDNA spawned a host of expression studies involving the expression of *CFTR* in many different cell types *in vitro* which suggested that CFTR was in fact a chloride channel (Rich *et al.* 1990., Kartner *et al.* 1991., Bear *et al.* 1991., Anderson *et al.* 1991., Tabcharani *et al.* 1991), as expression of *CFTR* in these cells coincided with the appearance of regulated chloride channel activity not previously detected.

This classification of CFTR as a chloride ion channel created more controversy. An ion channel has certain unique features, including ion selectivity, single channel conductance, and blocker sensitivity, by which it can be identified. However, previous studies had not all agreed on the 'signature' of the CFTR chloride channel. Some studies reported a 30pS-50pS outwardly rectifying chloride channel (ORCC), and others a much smaller 10pS chloride channel with a linear current-voltage relationship (reviewed in Frizzell and Cliff 1991). Three different chloride conductances had been identified in chloride-secreting epithelial cells (Cliff and Frizzell, 1990), and it was unclear which was defective in CF (the possibility of CFTR being responsible for more than one conductance deemed unlikely).

Definitive proof that CFTR was indeed a chloride channel came from the reconstitution studies of Bear *et al.* (1992), who studied the function of highly purified CFTR when incorporated into an artificial liposome bilayer. The chloride channel activity thus demonstrated ended not only speculation on the function of CFTR, but also the role of ORCC in CF, as the chloride channel exhibited the 10pS current and the linear current-voltage relationship of the non-rectifying channel. Therefore it was concluded that the ORCC probably had little to do with the CFTR chloride channel, and that earlier observations were 'unfortunate' (Higgins and Hyde, 1991).

This state of affairs remained until Egan *et al.* (1992) published evidence that not only are CFTR and the ORCC distinct chloride channels, but CFTR can regulate the ORCC. They arrived at this conclusion from the observation that the ORCC could

not be activated by protein kinase A in cells lacking CFTR, but this activation was restored on expression of *CFTR*. Subsequent studies on CF 'knockout' mice which lack CFTR (discussed later), reported the presence of ORCC in mouse epithelial cells indicating that CFTR and ORCC are indeed distinct chloride channels (Gabriel *et al.* 1993). The failure of protein kinase A to regulate ORCC in these epithelial cells lacking CFTR further supports the suggestion that they are regulated by CFTR.

Therefore, the basis of CF is defective chloride ion transport by the 10pS chloride ion channel which has a linear current-voltage relationship. A defect in this chloride ion channel may have many indirect but far-reaching effects, including defective regulation of other chloride ion channels.

1.1.3 Cloning The CF Gene

The CF gene was successfully cloned by a method known as 'positional cloning' or 'reverse genetics', which searches for linkage of a chromosomal region with the disease in question, and does not require a knowledge of the biochemical nature of the disease.

The first steps towards assignment of the CF gene to a particular chromosome arose when linkage was demonstrated to the enzyme paraoxonase (Eiberg *et al.* 1985), followed by linkage to the polymorphic marker DOCR1-917 (Tsui *et al.* 1985). Linkage to further polymorphic markers refined the critical region until two were identified which were separated by 10 kb, and were shown to flank the CF gene (Beaudet *et al.* 1986., Rommens *et al.* 1988). Using these markers as the starting points, Collins *et al.* (1987) used the chromosome jumping technique to move between the markers. Chromosome jumps 'filled in' with chromosome walking led them to a region which had maximum linkage disequilibrium with CF. Within this region a search was conducted for sequences which showed cross-species conservation indicative of functional DNA. This produced a candidate gene segment encoding a 6.2 kb transcript which had a short sequence of open reading frame preceded by a CpG island, and therefore some of the hallmarks of a gene (Rommens

et al. 1989). A search of a sweat gland cDNA library resulted in the identification of a 6.5 kb transcript with a pattern of expression that was restricted to exocrine tissues. Sequentially isolated cDNAs eventually led to a series of overlapping clones containing the entire coding region of the gene (Riordan *et al.* 1989). This gene was encoded by 250 kb of DNA, with 27 exons, and from which a protein of 1480 amino acids was translated. Expression was found to be higher in those tissues severely affected in CF, but no size difference was detected in the CF transcript. Confirmation that this was indeed the CF gene was demonstrated by the presence of mutations in this gene of individuals with CF which were absent in unaffected controls. This gene was validated as the CF gene as it fulfilled three essential criteria;

1. The gene was in the correct location and had no detectable recombination with the disease.
2. The nature of the gene product was consistent with the disease and had an appropriate pattern of expression.
3. Consistent mutations were found within this gene. The disequilibrium for one particular haplotype had suggested a common mutation, and the ΔF508 mutation was subsequently found to be present on 70% of CF chromosomes.

1.1.4 The Structure Of CFTR

Sequence analysis of the CF gene, named the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) predicted an encoded protein of 1480 amino acids with a molecular mass of 168 kd. The main structural features of this protein (Figure 1.1) deduced from amino acid sequence analysis were characteristic of a membrane bound protein due to the presence of two hydrophobic repeated motifs consisting of six α-helical sequences, typical of membrane spanning domains (MSD) (Riordan *et al.* 1989). Each membrane spanning domain is followed by a large hydrophilic region predicted to lie on the cytoplasmic side of the membrane, which has sequences resembling nucleotide binding domains (NBDs). These structural characteristics of CFTR have similarities to a family of ATP-dependent transport proteins called the

ATP binding cassette proteins (ABC). The most notable similarity being to the multiple drug resistant (MDR) proteins and the yeast STE6 protein, all of which import or export molecules across cell membranes in a process requiring ATP hydrolysis (Higgins 1989). There is one feature of CFTR which appears to be unique, that of the presence of a highly charged hydrophilic region lying in the middle of the protein in the cytoplasmic domain, the R domain. This region contains many polar residues and has multiple potential phosphorylation sites.

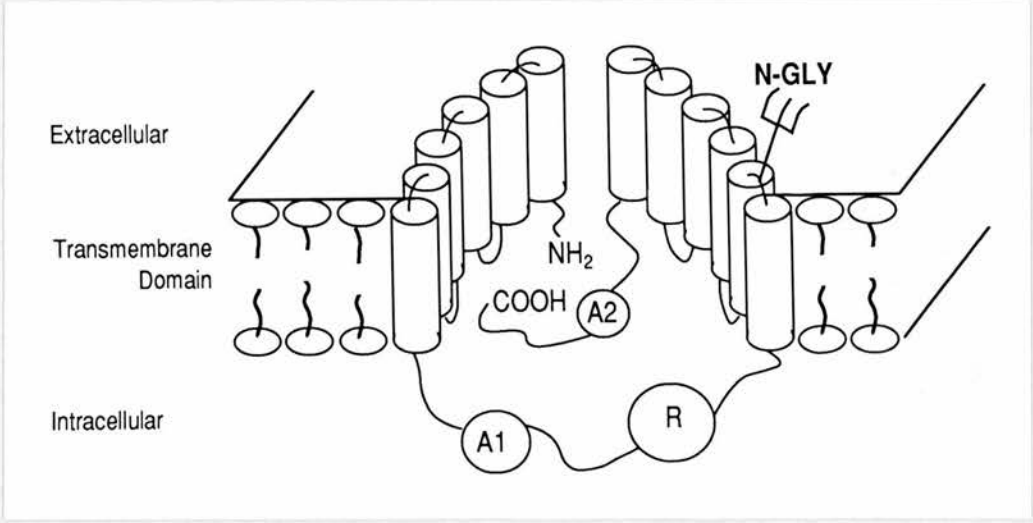


Figure 1.1 The proposed structure of CFTR
Schematic representation of the 3-dimensional structure of CFTR (taken from Dorin and Porteous 1991). The transmembrane domains are represented as cylindrical units, and N-GLY indicates a potential glycosylation site. The first nucleotide binding domain is denoted as A1, the second by A2, and the regulatory R domain is labelled R.

1.1.5 The Function Of CFTR

CFTR is a chloride channel. This is the conclusion that has been reached over the period since the CF gene was cloned. There has been much debate over the function of CFTR, and even now that there is general agreement on the role of CFTR as a chloride channel, the prospect of additional functions for CFTR has arisen.

If CFTR is a chloride channel, how does it carry out this function? It is thought that the transmembrane domains form a pore across the plasma membrane through which ions can pass. Alteration of the amino acids within this domain alters the anion selectivity of this channel (Anderson *et al.* 1992), indicating that the transmembrane domain both selects and conducts ions across the membrane.

The R domain, which lies between the two membrane spanning domains, is a feature of CFTR which is not shared by any known ABC protein. It is thought to regulate the chloride channel activity of CFTR, as its deletion generates chloride channels which are constitutively open (Rich *et al.* 1991). It has been proposed that CFTR channel activity is inhibited by a conformational change in the R domain, resulting in 'plugging' of the ion pore (Rich *et al.* 1991., Kartner *et al.* 1991., Cheng *et al.* 1991). The R domain is highly charged, and potential phosphorylation sites have been identified which if mutated, prevent cAMP dependent phosphorylation and activation of the chloride channel (Cheng *et al.* 1991). However, phosphorylation of the R domain alone does not activate the chloride channel, as binding of ATP by the nucleotide binding domains also appears to be required (Anderson *et al.* 1991).

The nucleotide binding domains (NBDs) which follow each membrane spanning domain, share extensive homology with the NBDs of other ABC transporter proteins (Higgins 1989). They contain Walker A and B sequences which interact with nucleotides (Walker *et al.* 1982), and have been shown to interact with ATP to regulate the opening and closing of CFTR chloride channels (Thomas *et al.* 1991., Anderson *et al.* 1991., Hartman *et al.* 1992). It has been found that there is a concentration of missense mutations in the first NBD, and a lack of a similar cluster in the second NBD, suggesting a difference in functional importance (Kerem *et al.* 1990., Cutting *et al.* 1990). Missense mutations are particularly informative as only one amino acid is altered in the mutant protein, and the observation that almost all missense mutations in *CFTR* occur in regions which are conserved between human, mouse and cow emphasise their functional significance (Dean *et al.* 1990., Diamond *et al.* 1991). This is supported by the expression studies of Gregory *et al.* (1991), which

studied the effect of CF mutations on the maturation and function of CFTR in Cos cells. They found that missense mutations in the first NBD almost always resulted in a lack of maturation or functional activity, but there was rarely any adverse effect seen with mutations in the second NBD. Anderson and Welsh (1992) described further evidence for the functional non-equivalency of NBD1 and NBD2. They demonstrated that ATP can bind to both NBDs, but that ADP binds to NBD2 only. This resulted in inhibition of chloride channel activity, which was abolished by mutations in NBD2. They suggested that ADP inhibits CFTR by competing with ATP at NBD2, and that the ATP : ADP ratio may be more important than absolute ATP concentration for regulating CFTR. Quinton and Reddy (1992) developed this hypothesis further by demonstrating that nonhydrolysable ATP analogues could stimulate CFTR in human sweat ducts following activation by cAMP, suggesting that binding rather than hydrolysis of ATP was required for CFTR chloride channel activity. Coupled with the observation that intracellular ATP concentrations must be around physiological concentrations for activation by this apparent non-hydolysable mechanism, this has led to the idea that by binding ATP and ADP, CFTR is either 'sensing' the energy level of the cell (ATP vs. ADP), or the cellular ATP concentration itself and it is this which regulates CFTR activity (Wine and Silverstein 1992).

Therefore it does appear that CFTR is a chloride channel, but does it have additional functions and are chloride ions the only molecules transported? The closely related human multidrug resistance P-glycoprotein has many structural similarities with CFTR, and belongs to the same group of ABC proteins. This protein has been found to be bifunctional, acting both as a pump which actively transports cytotoxic drugs out of the cell, and as a volume-regulated chloride channel (Valverde *et al.* 1992., Gill *et al.* 1992). By analogy, this raises the possibility of a dual function for CFTR. Hasegawa *et al.* (1992) demonstrated that CFTR can form a multifunctional aqueous channel capable of transporting anions, water and small solutes, and this was followed by evidence that fluid transport across airway epithelia was indeed defective in CF (Jiang *et al.* 1993). Tabcharani *et al.* (1993), have shown that CFTR functions as a multi-ion pore, transporting more than one anion simultaneously. In addition, they

demonstrated that some naturally occurring CF mutations associated with a mild phenotype, reduces the channel to a single ion pore with a concurrent 50% reduction in conductance. Therefore it appears that CFTR may have more roles than simply transporting chloride ions.

CFTR has also been implicated in other cellular functions. Bradbury *et al.* (1992) observed an absence of cAMP-dependent regulation of endocytosis and exocytosis in pancreatic cells derived from a CF patient, which was restored upon transfection with wildtype CFTR cDNA. This suggests a role for CFTR in regulation of membrane recycling which is an important way for cells to control both secretion and placement of proteins on their surface (Baringa, 1992). Barasch *et al.* (1991), reported that acidification of intracellular organelles such as the Golgi network, prelysosomes and endosomes was also defective in CF cells, and Lukacs *et al.* (1992) have reported that CFTR is present and functional in endosomes. The possibility that CFTR may be multifunctional may be important for understanding the relationship between this single gene defect and the many facets of this disease, as many of the clinical manifestations cannot be directly attributed to a simple defect in chloride ion transport. In particular, the abnormal composition of CF mucus which lacks sialic acid in the oligosaccharides, and the increased sulphation does not correlate directly with abnormal chloride ion transport (Baringa 1992., Richardson and Alton 1993), and might be more easily attributed to defective acidification of intracellular organelles.

1.1.6 CF Mutations

1.1.6.1 The spectrum of CF mutations

Following identification of the CF gene, a consortium was formed, The Cystic Fibrosis Genetic Analysis Consortium, consisting of 90 laboratories from 26 countries. This was set up to coordinate the accumulation of information on the identification of new mutations and their population frequencies, resulting in a large pool of information. To date, over 400 disease mutations have been identified (The Cystic Fibrosis Genetic Analysis Consortium 1994), almost half of which are missense

mutations. Non-sense, frameshift, and splice site mutations occur in roughly equal proportions, but few large deletions have been reported. No promoter mutations have been found, although this probably reflects a bias of mutation detection methods rather than an absence of such mutations.

Of all mutant CF chromosomes examined so far, 67% have been found to carry a three base pair (phenylalanine) deletion at position 508 (ΔF508). The frequency of this mutation varies in each population, and there is a clear Northwest to Southeast gradient of its frequency across Europe (De Braekeleer and Daigneault 1992), with the highest frequency reported in Copenhagen and Stockholm (CF Genetic Analysis Consortium 1990). The presence of the ΔF508 mutation at the high frequency of 1.4% (2% of chromosomes carry a CF mutation, 70% of which are the ΔF508 mutation) might be attributed to a founder effect, or alternatively, may imply an advantage to being a carrier, termed 'heterozygote advantage' (Knudson *et al.* 1967). This could be ascribed directly to the presence of a single CF mutation, or to an advantage conferred by closely linked loci (Wagner and Cavalli-Sforza 1975). It has been proposed that CF heterozygotes may have an advantage in surviving Cholera, which is characterised by excessive secretion of chloride and fluid in the intestine in response to the Cholera toxin (Rodman and Zamudio 1991). This was tested by Gabriel *et al.* (1994) using a mouse model for CF. They observed that mice homozygous for a null mutation in CFTR did not secrete fluid in response to Cholera toxin, but mice heterozygous for the null mutation secreted only 50% of the fluid volume lost by wildtype mice in response to the toxin. This reduced response in heterozygotes implies that they may be better able to survive the diarrhoea and resultant water loss associated with Cholera.

Different mutations can predominate in different populations, e.g. 87% of CF mutations in the Danish population are ΔF508 (The Cystic Fibrosis Genetic Analysis Consortium 1994), whereas only 22% are accounted for by ΔF508 in the Ashkenazi Jewish population where the W1282X mutation predominates (approximately 60% of CF chromosomes, Shoshani *et al.* 1992., Lerer *et al.* 1992). The isolated nature of

some communities contributes to the high frequency of certain CF mutations within these communities which is not seen in the general population, as demonstrated by the Hutterite community where the otherwise extremely rare M1101K mutation accounts for 69% of CF chromosomes (Zielenski *et al.* 1993).

1.1.6.2 Screening for CF mutations

The large number of CF mutations and their variant frequencies makes detection of 100% of mutations within a population impossible, and therefore presents a dilemma for advocates of population screening programmes. Initially screening could detect only 70% of CF carriers thus identifying only 50% of at risk couples, a level that many found unacceptably low (Roberts 1990., Beaudet 1990., Gilbert 1989). Significant improvements have been made in the detection of CF mutations in various populations (Shoshani *et al.* 1992., Super and Schwarz 1992., Cheadle *et al.* 1992), with the highest rate of detection of 98% reported by Ferec *et al.* (1992) in a Celtic population.

1.1.6.3 Correlation of genotype with phenotype

There is a great range of symptoms and degrees of severity associated with CF, and their correlation with a particular genotype has often been attempted. The ability to predict the course and severity of the disease for each individual would be of enormous value both for informed decision-making following a positive prenatal diagnosis for CF, and for counselling and care of an affected individual. A recent report from The Cystic Fibrosis Genotype-Phenotype Consortium (1993) attempted to correlate the phenotypes of approximately 62% of CF patients with their genotype. They concluded that 'the only prognostic value of genotypic information is for the prediction of pancreatic status'. The association that has become apparent is the correlation of so called 'severe' alleles and pancreatic insufficiency (PI) (Kristidis *et al.* 1992). One such severe allele is ΔF508 which, in the homozygous state, has been frequently correlated with pancreatic insufficiency (Kerem *et al.* 1990., Santis *et al.* 1990., Pignatti 1991). When CF individuals are homozygous for a severe mutation or have two different severe mutations, then they are usually PI. Individuals with a

severe and a 'mild' mutation e.g. R117H, are usually pancreatic sufficient (Johansen *et al.* 1991., Kristidis *et al.* 1992., Osbourne *et al.* 1992., Ferec *et al.* 1993). The mild mutation being dominant over the severe mutation. Four mutations have been designated severe, these are ΔF508, G542X, R553X, and W1282X, and prenatal and prognostic counselling for homozygotes and compound heterozygotes should advise of longterm PI (CF Genotype-Phenotype Consortium, 1993). A reduced risk of meconium ileus has been reported for the G551D mutation (Hamosh *et al.* 1992). Homozygosity for ΔF508 has also been associated with an earlier age of diagnosis (Kerem *et al.* 1990., Hamosh *et al.* 1992), but has not been found to correlate with severity of pulmonary disease (Santamaria *et al.* 1992., Santis *et al.* 1990). The conclusion of the CF Genotype-Phenotype Consortium is that severity and course of pulmonary disease are not predicted by genotype, and they suggest that factors other than CF genotype affect the pulmonary phenotype. Kiesewetter *et al.* (1993) reported that the variation of the phenotype of CF patients carrying at least one R117H mutation in association with another CF mutation correlated with the chromosomal background. Correlation between CF mutations and other aspects of CF have been attempted but no clear picture has emerged. Dodge (1991) cautions that the contribution of non-genetic factors must not be overlooked, especially as identical twins with different health status' have been reported, and that there will always be great variability in exposure to infection.

1.1.6.4 The effect of CF mutations on CFTR function

The consequences of the ΔF508 mutation on CFTR function have been studied extensively. Cheng *et al.* (1990) studied the expression of mutant ΔF508 *CFTR* cDNA in Cos cells and observed that mature CFTR was absent in these cells. Instead, an incompletely glycosylated form was detected which they suggested was a result of incomplete CFTR processing in the endoplasmic reticulum (ER), resulting in its degradation and therefore absence at the correct cellular location. The ΔF508 allele has been shown to be expressed at the same level as the normal allele in the respiratory tract (Trapnell *et al.* 1991) indicating that transcription of mutant *CFTR* is not defective. Further study into the processing and intracellular location of ΔF508

CFTR produced conflicting results. Studies examining the maturation and function of mutant *CFTR* expressed in transfected cells *in vitro* (Gregory *et al.* 1991., Dalemans *et al.* 1992), agreed with those of Cheng *et al.* (1990). Nevertheless, other studies detected correct localisation of ΔF508 CFTR in the plasma membrane (Sarkadi *et al.* 1992), and chloride channel activity when ΔF508 cDNA was expressed in *Xenopus* oocytes (Drumm *et al.* 1991), Vero cells (Dalemans *et al.* 1991), and Sf9 insect cells (Bear *et al.* 1992). A subsequent publication was able to reconcile these differences by demonstrating that the processing of mutant CFTR is temperature sensitive (Denning *et al.* 1992). They showed that processing of ΔF508 CFTR reverts to wild type as the temperature is lowered, thereby explaining the differences seen in processing and localisation in experiments using oocytes and insect cells which are typically conducted at lower temperatures. The anomalous results of the experiments using Vero cells were obtained in a very high expression system and therefore might be a result of some protein escaping ER 'quality control mechanisms' and reaching the plasma membrane where it could generate some chloride channel activity. This was confirmed by an important observation that the ΔF508 mutation was also mislocalised in sweat gland biopsies taken from CF patients (Kartner *et al.* 1992). This led to speculation that the basis of defective CFTR channel activity of ΔF508 mutants might be entirely attributable to its mislocalisation, which if overcome, might have sufficient activity to correct the defect. This was confirmed by Li *et al.* (1993) who demonstrated that ΔF508 mutant CFTR exhibited phosphorylation-regulated chloride channel activity similar to that of wildtype CFTR when reconstituted into lipid bilayers. Yang *et al.* (1993) followed this by showing that incubation of L cells transfected with ΔF508 cDNA at 26°C for 48 hours elicited an electrophysiological response which was absent in cells incubated at the higher temperature of 37°C. This observation is of great importance for the treatment of CF, as it implies that if ΔF508 CFTR could be relocated to the plasma membrane, the mutant protein should exhibit enough chloride channel activity to ameliorate or even correct the CF defect.

Although defective processing is also observed with other CF mutations, this is not the sole mechanism by which CF-causing mutations exert an effect. Welsh and Smith

(1993) have divided CF mutations into four classes based upon their effect on CFTR biosynthesis and function (table 1.1, and figure 1.2):

Class I includes all CF mutations which result in defective protein production. This might be due to the presence of premature termination signals, the production of unstable mRNA, or severely truncated or aberrant protein which is unstable and quickly degraded. All mutations of this type have the net result that no detectable CFTR protein is produced.

Class II mutations are those which fail to progress through the biosynthetic pathway and traffic to the correct cellular location, and includes the ΔF508 mutation. The basis of the defect caused by this class of mutations is the absence of CFTR at the plasma membrane.

Class III mutations are processed correctly, and do reach the plasma membrane, but exhibit defective regulation of the chloride channel activity. These mutations have been found to be located in the nucleotide binding domains, and NBD1 in particular, which is consistent with its proposed greater participation in the regulation of CFTR than NBD2.

Class IV mutations affect the conductance of CFTR and have found to be located in MSD1. These are mutations which do not affect the processing or localisation of the mutant protein, but reduce the CFTR current.

There appears to be some correlation between the mechanism by which CF mutations disrupt CFTR function and the severity of pancreatic disease (Welsh and Smith 1993). Sheppard *et al.* (1993) have shown that mutant CFTR containing mutations associated with mild pancreatic disease retains significant apical chloride channel activity when expressed in epithelial cells. This residual chloride channel activity would explain the dominant nature of mild mutations when combined with a severe allele in compound heterozygotes.

Table 1.1 Classes of *CFTR* Mutations That Cause CF

CLASS	DEFECT	EXAMPLES	DOMAIN	FREQUENCY	CLINICAL STATUS
I	NO PROTEIN				
	Nonsense mutations	G542X	NBD1	3.4	PI
	Frameshift mutations	3905 insT	NBD2	2.1	PI
	Splice site mutations	621+G→T	MSD1	1.3	PI
II	PROCESSING	ΔI506/7	NBD1	0.5	PI
		ΔF508	NBD1	67.2	PI
		S549I	NBD1	RARE	
		S549R	NBD1	0.3	PI
		A559T	NBD1	RARE	
		N1303K	NBD2	1.8	PI
III	REGULATION	G551D	NBD1	2.4	PI
		G551S	NBD1	RARE	PS
		G1244E	NBD2	RARE	PI
		S1255P	NBD2	RARE	PI
		G1349D	NBD2	RARE	PI
IV	CONDUCTION	R117H	MSD1	0.8	PS
		R334W	MSD1	0.4	PS
		R347P	MSD1	0.5	PS

(Taken from Welsh and Smith 1993).
Frequency of mutations is expressed as percentage of all CF mutations.
NBD refers to nucleotide binding domain, MSD membrane spanning domain, PI pancreatic insufficiency, and PS pancreatic sufficiency.

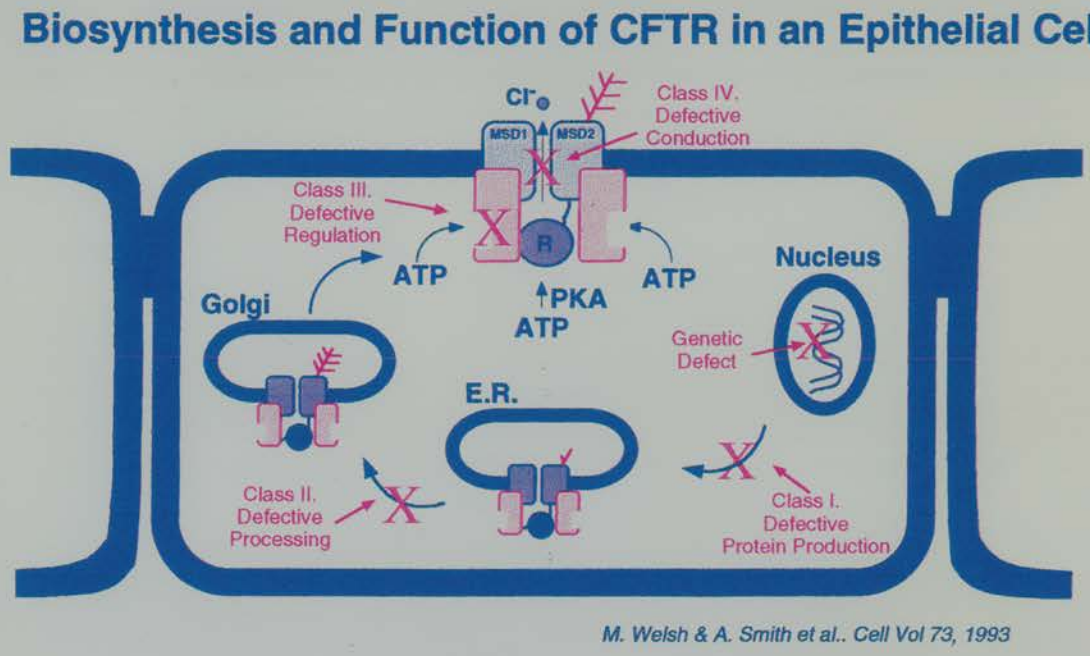


Figure 1.2 The biosynthesis and function of CFTR in an epithelial cell.

Glycosylation of the protein is indicated by the branched structure, MSD1 and MSD2 refer to the membrane spanning domains, and R to the regulatory R domain. The pink regions flanking the R domain are the nucleotide binding domains. The cross represents the block in the biosynthesis or function of CFTR.

1.1.7 The Expression Of CFTR

Cell-specific expression of *CFTR* was studied by Trezise and Buchwald (1991) using labelled antisense RNA probes for *in situ* analysis on rat tissue sections. They found that *CFTR* was specifically expressed in the ductal cells of the pancreas and salivary glands, and that there were decreasing gradients of expression in the intestine along the crypt-villus, proximal-distal axes. *CFTR* expression was also detected at low levels in the mucosa and submucosa of the bronchi and bronchioles of the lung, with the highest expression being detected in the sub-mucosal glands. Expression in the testis was found to be regulated during the cycle of the seminiferous epithelia, and implies that CFTR plays a role in spermatogenesis thereby contributing to CF infertility.

The localisation of CFTR has been studied immunocytochemically using antibodies raised against CFTR. Great difficulty has been experienced in raising antibodies to CFTR and in achieving a uniform pattern of expression. Crawford *et al.* (1991) found abundant CFTR in epithelial cells including those lining the sweat ducts, small pancreatic ducts, and intestinal crypts. The level of CFTR in lung epithelia was again found to be surprisingly low, but was abundant in kidney tubule epithelia. The protein appeared to be restricted to the apical membranes of cells, a feature also described by Cohn *et al.* 1991., Denning *et al.* 1992., and Zeitlin *et al.* 1992., which, along with tissue distribution is consistent with its proposed role as a chloride channel.

Expression of *CFTR* during foetal development has also been studied using *in situ* techniques to detect mRNA (Tizzano *et al.* 1993., Trezise *et al.* 1993). This demonstrated expression of *CFTR* in all major organs known to be affected in CF, with an expression pattern which broadly matched that of the adult. However, these studies also identified some interesting differences between adult and foetal *CFTR* expression. The level of *CFTR* expression in the foetal lung epithelium was higher than that seen in adults, and led to the suggestion that *CFTR* might fulfil a different role in the aqueous environment of the foetal lung (Tizzano *et al.* 1993). In addition, there was a conspicuous lack of expression in the foetal sub-mucosal glands, which is

the site of highest expression in the adult. Expression of *CFTR* in reproductive tissues varied between males and females. No expression was detected in the first and second trimester of foetal development in females, however significant expression was apparent in uterine, cervix and fallopian tube epithelia in the third trimester. In males, a low level of expression was detected in the epididymis epithelia at all stages. However, no expression of *CFTR* was detected in ovary or testis at any stage of foetal development.

Although *CFTR* has been shown to be expressed in all tissues affected by the disease process of CF, the levels of expression do not appear to correlate with the degree of pathology. However, there is no reason to presume that the major sites of disease pathology would exhibit the highest levels of expression, as each tissue and organ might have differential expression requirements (Porteous and Dorin 1993). Alternatively, the variation in disease pathology might be more a reflection of varying tissue vulnerability to defective chloride transport, or the consequent luminar blockage.

1.1.8 Cross Species Analysis

Cross species analysis of *CFTR* sequences has demonstrated a high level of conservation indicating that CFTR has been highly conserved throughout evolution. Murine CFTR has an overall identity to human of 78.3% at the amino acid level, which increases to 89.6% if conservative changes are included (Yorifuji *et al.* 1991., Kelley *et al.* 1992). Analysis of rat *CFTR* indicates a nucleotide sequence identity of 80.5%, and an amino acid sequence identity of 75.5% (Trezise *et al.* 1992), which is similar to that of 77.4% reported for *Xenopus laevis* at the amino acid level, and 72% for dogfish CFTR (Marshall *et al.* 1991). Analysis and comparison of many diverse species such as mouse, rat, Guinea pig, rabbit, sheep, cow, pig, dogfish, marmoset, and *Xenopus laevis,* has identified that the highest levels of sequence conservation are always seen in the regions encoding the first membrane spanning domain and the first nucleotide binding domain (Marshall *et al.* 1991., Gasparini *et al.* 1991., Yorifuji *et al.* 1991., Diamond *et al.* 1991., Kelley *et al.* 1992., Trezise *et al.* 1992., and Tucker

et al. 1992). The region which is consistently least conserved across species is the R domain. This endorses the predictions of CFTR functional assays which suggest that MSD1 and NBD1 play an important role in regulation of CFTR chloride channel activity through specific interactions, whereas regulation of CFTR by the R domain occurs through a 'steric' non-specific mechanism. In addition it has been observed that in dogfish, mouse, and cow, most of the missense CF-causing mutations in humans occur at conserved sites. This high level of cross species conservation, and in particular conservation of the sites of human missense mutations, augurs well for studying the disease in an animal model.

1.2 MODELLING HUMAN GENETIC DISEASE IN AN ANIMAL SYSTEM

1.2.1 The Value Of Animal Models

Animals have been used extensively in science to improve our knowledge in many different ways. By our study of animals we have widened our understanding of many natural biological processes and extrapolated from these to the mechanisms of disease. To increase the usefulness of these studies in furthering our understanding of disease etiology, it would be desirable to obtain animals which exhibit many if not all of the features of a human genetic disease. Study of some diseases in man is often limited by the restricted access to human tissue which often means that very few of the physiological changes underlying many diseases can be ascertained. Tissues are often only available at autopsy by which time the destructive final stages of the disease can leave few clues to its etiology. These same restrictions however do not hold for animals, and access can be gained to organs and tissues at all stages of the disease process. Furthermore, the study of human genetic disorders is frequently hampered by the contribution to the disease phenotype of many features which can not be fully identified and evaluated. These might be the influence of environmental factors, or genetic background, or even compliance with prescribed therapy. Use of laboratory animals which have been inbred to isogenicity should eliminate any variation due to genetic background, and their maintenance in a controlled environment should allow

any environmental influence to be stabilised. Therefore the many variables of human disease study can be removed or controlled by use of laboratory animals. In addition, the typically shorter life cycle and increased fecundity of many animals compared to humans would allow a greater number of affected individuals to be studied thereby attaching a greater level of significance to any conclusions which might be drawn. Use can be made of the techniques for culturing organs, tissues, and cells *in vitro*, however these systems can never be directly analogous to the situation *in vivo* and are often subject to artefactual effects.

Many useful animal models of human genetic disease possess mutations which have arisen spontaneously (Kolberg 1992), however this relies on recognition of the mutant animal and the phenotype which parallels a human condition. The generation of these spontaneous mutations are likely to be relatively rare events, and models are only likely to be found for diseases with very characteristic and easily identifiable phenotypes. Consequently few recessive lethal mutations have been detected this way as these are by their nature difficult to detect. An alternative strategy is to increase the probability of mutation in the gene of interest by exposing animals to known mutagens. This again relies on screening a very large number of animals, identification of a characteristic phenotype and relating this to the human disease in question. This too has the problem that the nature of the causative mutation is often difficult to define, and can limit the usefulness of such a model.

1.2.2 Choice Of Animal And Technology

When choosing an animal in which to model a human disease it is important to consider the likelihood that the disease in this animal will be analogous to that of the human condition. The biology of the chosen species should be well characterised and indicate that it has a similar metabolism to man. To this end the comparability of the tissue architecture and the possibility of compensatory mechanisms must be taken into account. Moreover, if the genome of an experimental animal is to be manipulated to model a human genetic disease, use of a species in which the genetics have been well characterised would be expedient. The gene and/or the region to be manipulated

should show a high degree of sequence and functional conservation to that of the human gene implying the similarity of a mutant phenotype.

However a choice based on the above criteria must also be balanced against other considerations such as the convenience and cost of housing different species, and the ease of handling. Primates are an obvious choice due to their close evolutionary relationship to man, however many primate species are endangered, and they are typically expensive to maintain and difficult to work with. Their low fecundity means that the period taken to breed mutant animals in significant numbers could be unacceptably long, and indeed some maintain that studies of primates have little advantage over humans (Lathe and Mullins 1993). Rodents are frequently the animals of choice as they are relatively cheap to maintain, have comparable biology, and high fecundity. For this reason, both rodent biology and genetics have been studied extensively.

Other considerations include the availability of suitable technology for the species of choice. The choice of an appropriate technology would depend upon the nature of the genetic defect in man. For dominant mutations, technologies can be utilised which simply add a copy of the mutant allele to the genome. By introducing the human DNA (transgene) into the germ line of an animal, the mutant allele can be passed on to many animals simply by breeding. The vector can be designed to give the transgene specific features such as tissue specific expression (or lack of), and also to vary the level of expression. A transgenic mouse model for Retinoblastoma was produced in this manner by expressing a viral oncogene in the retina of mice (Windle *et al.* 1990). The transgene is usually introduced into an early stage embryo to maximise the chances of it integrating into the genome of cells destined to become germ cells. The most common method is to use microinjection to introduce the DNA into a fertilised egg, which typically results in the stable integration of a head-to-tail array of the transgene at a random site in the host genome. This occurs at a sufficiently high frequency to establish it as a useful technique but has the disadvantage that the site and nature of the integration cannot be predicted. There

can be a huge variation in copy number of the transgene which affects its level of expression, and the random nature of the transgene integration means that there is also the possibility that the integration site influences the resulting phenotype. Some sites of integration can abolish transgene expression, and integration into an essential gene can result in a lethal phenotype. Therefore although this method is suitable for modelling some diseases in an animal, the imprecise nature of the transgene insertion makes it unsuitable for others. In addition, this technology is only suitable for addition of genetic material and not of use for studying loss of function mutations.

Many human genetic diseases however, are not a result of a dominant mutation and therefore cannot be created by addition of a mutant allele. This thesis is concerned with modelling one such human genetic disease, cystic fibrosis, in an animal system. This is a recessive genetic disease which along with other diseases of this type arises through mutation of both alleles of a gene. To produce an animal model for this disease, both alleles of the gene must be selectively mutated resulting in loss of function of the resultant gene product. There is currently only one method available for achieving this precisely, gene targeting in embryonic stem (ES) cells. The manipulation of DNA by gene targeting can be applied to most if not all animals, but the technology to create an animal which bears this targeted alteration is only possible through the use of ES cells. Furthermore, it is only currently possible to generate animals from murine ES cells. There have been reports of ES cell isolation for a few other species such as rat (Iannaccone *et al.* 1994), rabbit (Graves and Moreadith 1993) and mink (Sukoyan *et al.* 1993), however no animal entirely derived from these ES cells has as yet been reported. Despite this lack of choice, the mouse should still be a good choice in which to model cystic fibrosis as it fits many of the criteria for choosing an animal model. Mice are both cheap and easy to house, and can produce many large litters in a relatively short period. Their biology is comparable to that of man, and their genetics have been studied extensively. Sequence analysis has shown that there is a high degree of conservation between murine and human CFTR with an overall identity at the amino acid level of 78.3%, which rises to 89.6% if conservative changes are included (Yorifuji *et al.* 1991., Kelley *et al.* 1992). In addition, almost

all of the CF-causing mutations reported in man occur at sites which are conserved in the mouse. This suggests that mutation of any of these conserved regions in CFTR is likely to be of equal functional significance to both man and mouse. Smith *et al.* (1992) have shown that wildtype mice have similar ion transport properties to those of normal human subjects, implying that a genetic mouse model for CF would be likely to display a similar clinical phenotype.

1.2.3 Embryonal Stem Cell Technology

Embryonal stem (ES) cells are cells which have been isolated from the inner cell mass of a 3.5 day mouse blastocyst (Evans and Kaufman, 1981., Martin 1981). When maintained under stringent conditions, ES cells retain the normal pluripotent phenotype of the embryonal cells from which they originated, over many passages and following genetic manipulation. On return to the blastocyst environment, pluripotent ES cells can recommence their embryonic development and contribute to the tissues of the resulting offspring including the germ cells (Bradley *et al.* 1984). This unique property of ES cells enables them to be maintained *in vitro* where they are accessible to genetic manipulation, and then returned to a blastocyst to generate mice which can be bred to obtain animals which are homozygous for the introduced alteration. The animals resulting directly from such manipulated blastocysts usually have a genetic contribution from both the host blastocyst and the reimplanted ES cells, and are said to be chimaeric. The ES cell derived component can vary enormously between different tissues of the resulting chimaera, but contribution to the germ cells is required for the mutation to be passed through the germ line to generate mice which are homozygous for the planned alteration. Originally, embryonal carcinoma (EC) cells were used, which are cells isolated from embryonal teratocarcinomas (Solter *et al.* 1970). However these were found to have a restricted pattern of differentiation and a poor ability to colonise the tissues when introduced into a blastocyst (Mintz and Illmensee 1975).

Due to their pluripotent nature, ES cells are particularly problematic to culture *in vitro*. Many known and unknown factors can stimulate them to differentiate,

especially any stress encountered in their culture conditions. Differentiation is undesirable as it is irreversible and renders cells incapable of further growth and blastocyst colonisation. Originally, ES cells were cultured on feeder layers of growth-arrested fibroblasts (Martin, 1975). Differentiation of ES cells was mostly inhibited with this system, and in some part appeared to be mediated by a diffusible factor, as medium alone conditioned by feeder cells was found to also inhibit differentiation (Smith and Hooper 1983). This diffusible factor was also found to be present in medium conditioned by Buffalo rat liver cells (Smith and Hooper 1987). Further analysis of this factor (named Differentiation Inhibitory Activity-DIA) led to its characterisation and found it to be identical to a known myeloid leukaemia inhibitory factor (LIF) (Williams *et al.* 1988., Smith *et al.* 1988). A recombinant form of LIF/DIA was produced by transient expression in Cos-7 cells which could support the maintenance of ES cells in the absence of feeder layers (Smith *et al.* 1988).

Following genetic manipulation *in vitro*, pluripotent ES cells carrying a desired alteration can be used to generate mutant mice by the scheme presented in figure 1.3. The first step is to return the altered ES cells back to the blastocoel environment. The most commonly used method utilises microinjection to deliver the cells to the blastocoel where they become incorporated into the inner cell mass (Bradley 1987). This technique is technically demanding and utilises expensive equipment. Recently a less technically demanding method of introducing cells into the blastocyst has been developed which takes advantage of the natural aggregation properties of preimplantation embryos (Wood *et al.* 1993 ., Nagy and Rossant 1993). With this method, clumps of eight to twelve ES cells are cultured in close proximity to eight cell stage morulae which have had the zona pellucida enzymatically removed. The ES cells become internalised and incorporated into the morulae which go on to from blastocysts *in vitro*. Due to the simplicity of this technique, a large number of chimaeras can be generated in a short period of time. Another new development in this field, is the possibility to generate mice which are completely derived from ES cells (Nagy and Joyner 1993). This technique uses tetraploid embryos to provide the extra-embryonic structures required by the embryo, which ES cells themselves are

unable to generate. This method has the advantage that dominant mutations, or mutations which are already homozygous in the ES cells, do not need to be passed through the germ line of chimeras, however it also has the major disadvantage that mice derived from ES cells in this manner have to be delivered by Caesarean section as they usually die shortly after birth. In addition, only a few of the ES cell lines which generate germ line chimaeras by blastocyst injection can generate completely ES cell derived embryos by this technique.

Once the manipulated ES cells have been introduced into blastocysts by whatever method, they are implanted into the uterus of pseudopregnant host females where they go on to complete normal embryonic development. Pigmentation markers are often used to give some visual representation of the contribution of the ES cells to the resulting chimaeras. A common combination is to introduce ES cells carrying non-pigmented or reduced pigment markers into blastocysts from a more heavily pigmented strain. Any contribution of the ES cells to the melanocytes of the skin will be seen as a lightening of the coat colour, and contribution to eye seen as loss of pigmentation. Although this provides some indication of ES cell contribution it does not give any information as to the contribution to the germ cells, which is the most important feature if the mutation is to be passed through the germ line. However a chimaera with a high degree of ES cell coat colour contribution is also likely to have some contribution to the germ cells. Isoenzyme analysis can also be utilised, but again does not furnish information regarding germ cell contribution. Another observation of ES cell contribution can be obtained from the sex distortion ratio. Most ES cell lines in use were derived from male embryos and therefore have an XY phenotype. Male ES cells are usually preferred, as female ES cells are prone to losing an X chromosome during culture *in vitro* to become an XO genotype. In addition they have the advantage that male ES cell contribution to the developing gonadal ridge results in expression of the Y chromosome-specific *SRY* sex determining gene. If expressed at a sufficient level, this dosage-dependent protein can result in the sex conversion of a female embryo to male (Koopman *et al.* 1991). The resulting male animal will only be able to make sperm from the male ES cell contribution as female

XX cells are unable to undergo spermatogenesis. Therefore the chances of a chimaera passing the ES cell mutation through the germ line is much greater in these animals which have no contribution to the germ line from the host blastocyst. Also any distortion of the expected 50:50 ratio of male to female chimaeras gives some indication that sex conversion has occurred, indicating ES cell germ line contribution.

By breeding chimaeras with ES cell contribution, it should be possible to obtain animals which have inherited the ES cell mutation which has been passed through the germ line. These animals will be heterozygous for this mutation, and by breeding these together it should be possible to generate animals homozygous for the mutation. Any absence of such homozygotes would indicate an embryonic lethal mutation.

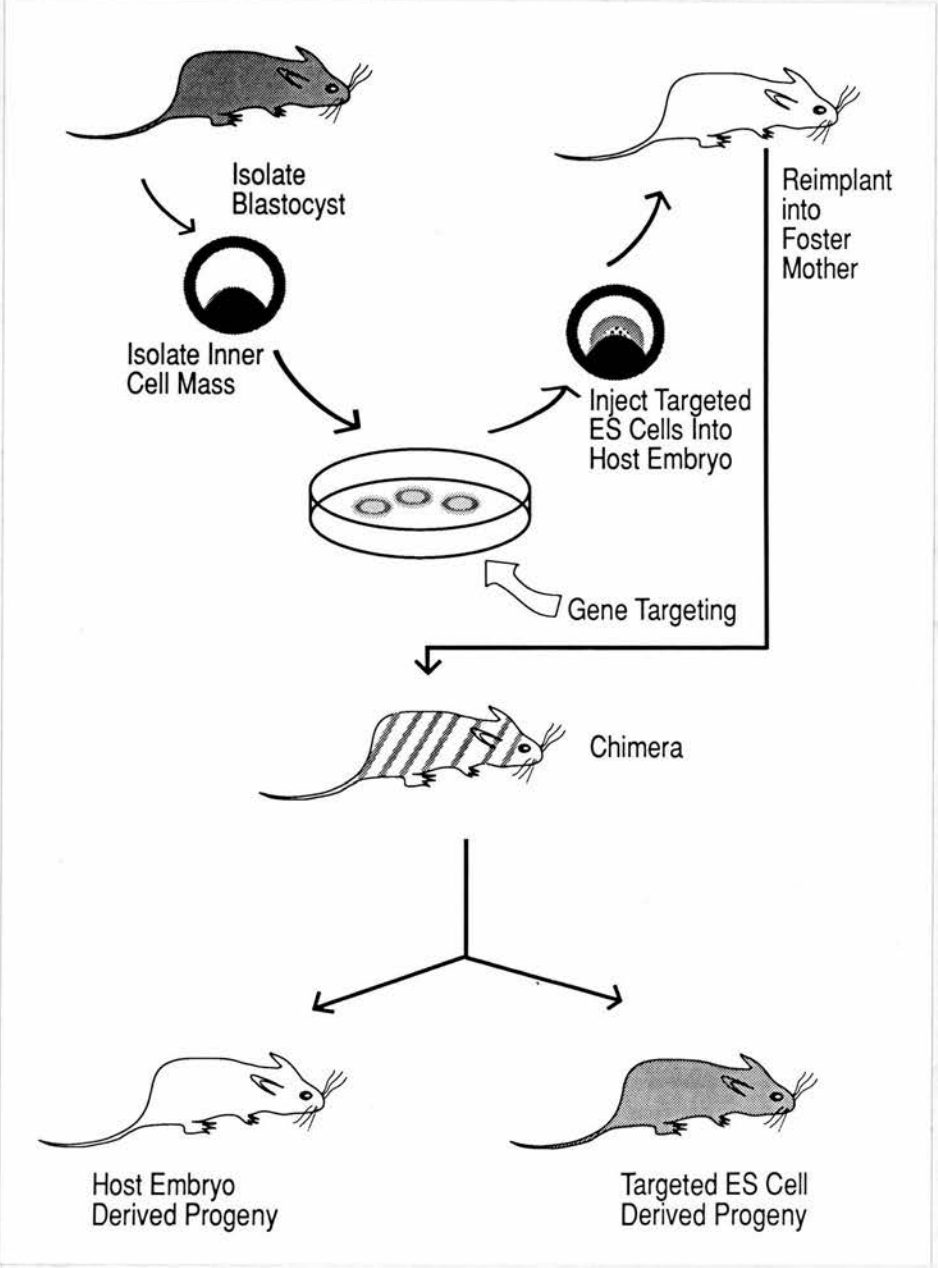


Figure 1.3 Generating transgenic mice from ES cells.

1.2.4 Gene Targeting

In order to make use of ES cell technology, a method must be available to mutate the gene of interest in ES cells *in vitro*. The most commonly used method is gene targeting and is dependent upon a cells intrinsic ability to mediate recombination between homologous DNA sequences. This process was first demonstrated in the yeast *Saccharomyces cerevisiae*, when a *LEU2* deficient mutant was transformed by integration of a plasmid carrying wildtype *LEU2* sequences. This plasmid was shown to have integrated into the target sequence through a single homologous reciprocal exchange (Hinnen *et al.* 1978). Transforming DNA was subsequently found to integrate into the genome of yeast exclusively through recombination with homologous sequences (termed homologous recombination) and was used as an efficient method of replacing or disrupting many endogenous genes (Scherer and Davis 1979., Shortle *et al.* 1982). Orr-Weaver *et al.* (1981) showed that DNA ends are highly recombinogenic, and therefore the frequency of homologous recombination could be stimulated by introducing a double strand break in the transforming DNA. In addition, integration of a plasmid carrying a gap in the region of homology usually resulted in repair of the gap using the chromosome sequences as the template. These observations led Szostak *et al.* (1983) to suggest that homologous recombination was mediated through direct interaction of the free DNA ends of the homologous sequences, and they proposed the 'double-strand-break repair' model to account for the mechanism of integration through homologous recombination (figure 1.4). This model proposes that the double strand break in the homologous sequence is enlarged through the action of endonucleases to expose 3' single-stranded termini. One of these free 3' ends then invades a homologous region of the chromosome DNA, displacing a small D loop. This D loop is enlarged by repair synthesis until sequences to the other 3' end are exposed allowing this free end to also anneal. Repair synthesis completes the gap repair process, and branch migration results in the formation of two Holliday junctions. Depending on the resolution of the Holliday junction, gap repair will be followed by a non-crossover event, or by a crossover event and integration of the exogenous DNA.

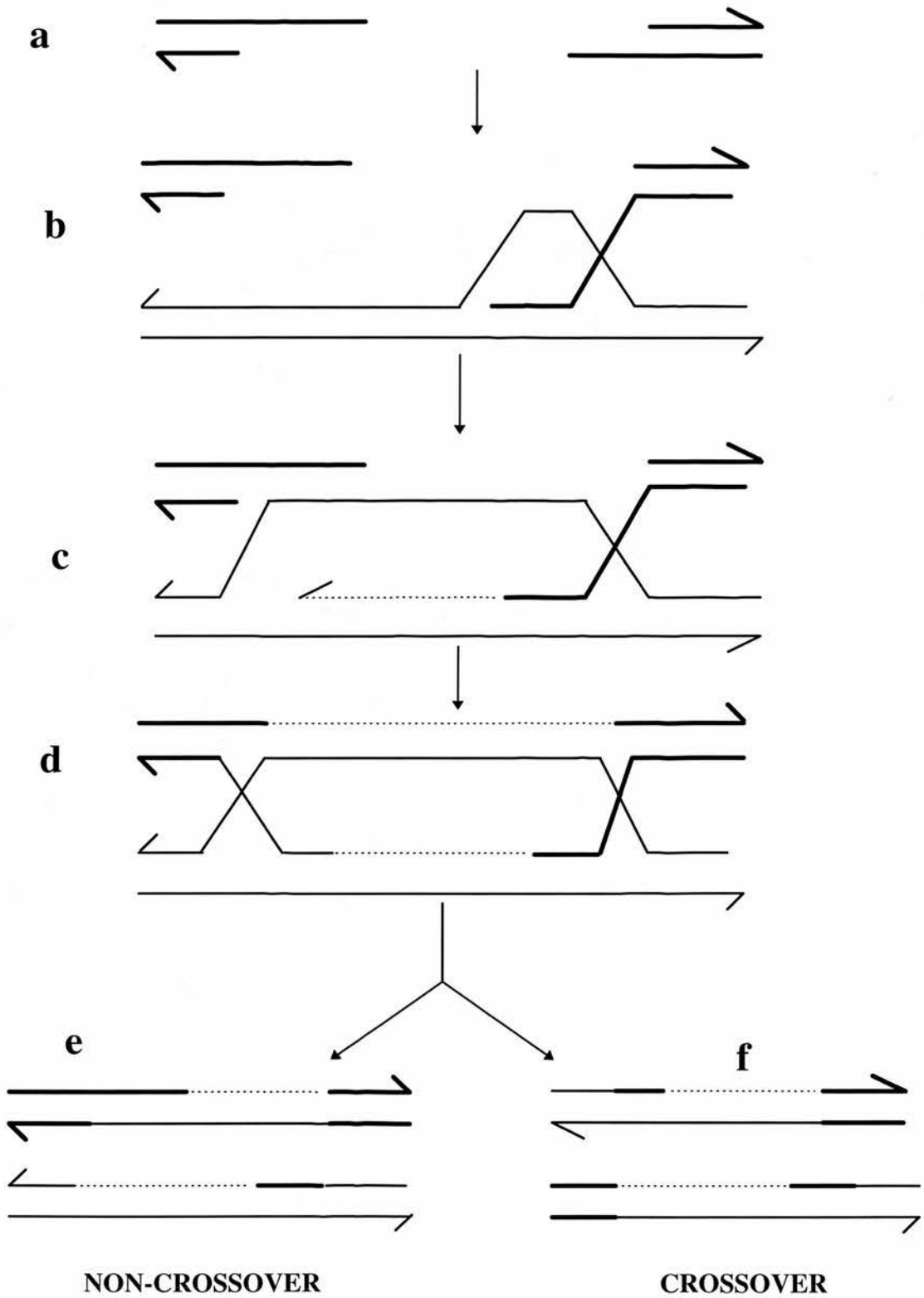


Figure 1.4 The double-strand-break repair model.
(a) Double strand cut in DNA. (b) Invasion of 3'end displacing D loop. (c) Enlargement of D loop. (d) Repair syntheis from other 3' end completes gap repair. Branch migration results in formation of two Holliday junctions. Resolution of junctions leads to either (e) non-crossover, or (f) crossover configurations. Taken from Szostak *et al.* (1983).

Various experiments have demonstrated that mammalian cells are also capable of mediating homologous recombination between newly introduced DNA molecules and chromosomal regions of homology. Lin *et al.* (1984) demonstrated that the TK sensitivity of mouse L cells containing a mutant *tk* gene could be restored through homologous recombination with exogenous DNA bearing a different mutation. A similar experiment was conducted by Thomas *et al.* (1986) who demonstrated the correction of defective neomycin genes in mouse L cells by homologous recombination. Demonstration of this phenomenon at an endogenous locus was shown by Smithies *et al.* (1985) by targeting the β globin gene in erythroleukaemia cells. As in yeast, the frequency of homologous recombination was found to be increased if the transforming DNA was linear (Folger *et al.* 1982., Lin *et al.* 1984., Valancius and Smithies 1991). Likewise any gaps in the region of homology carried by the incoming DNA are corrected using the chromosome as the template (Valancius and Smithies 1991). Analysis of the sites of homologous recombination has shown that DNA integrates into the target site in a very precise manner whilst retaining the fidelity of the target sequence (Zheng *et al.* 1991). However unlike the situation in yeast, mammalian cells have a greater propensity for non-homologous recombination, which is the predominant event (Roth *et al.* 1985).

1.2.4.1 Targeting vectors

Gene targeting has developed into a widely used method of mutating DNA at a precise chromosomal location in a predetermined manner. Improvements in our understanding of the mechanism of homologous recombination have led to refinements in vector design and the type of mutation which can be created. There are two basic types of targeting vector which differ in their mechanism of integration, and by careful design these vectors can be used to introduce a whole range of different mutations. Both vector types incorporate basic features such as regions of homology to the target site to mediate homologous recombination, which are carried on a plasmid backbone. As transfection efficiency is often low, a positive selectable marker is commonly included to facilitate the isolation of transfected cells.

One of these types of targeting vectors is an insertional or (O) type vector. These are linearised within the region of homology to the target site and integrate via a single reciprocal recombination event stimulated by free ends as predicted by the double-strand-break repair hypothesis (Figure 1.5). The entire targeting vector is inserted into the target site through this mechanism of integration and results in a duplication of the homologous sequences separated by the plasmid sequences. As the sequences of the target site are simply disrupted with this type of integration, there is the possibility that wildtype mRNA might be generated as a consequence of alternative splicing and exon skipping. This phenomenon has been observed at a low level in mice in which targeting was predicted to create a null allele at the N-*myc* locus (Moens *et al.* 1992) and at the *Cftr* locus (Dorin *et al.* 1994), although absolute null mutations have been generated using insertional vectors (Deng *et al.* 1992., O'Neal *et al.* 1993). The frequency of homologous recombination has been reported to be higher with this type of targeting vector, than with the alternative replacement type vector (Hasty *et al.* 1991., Dickinson *et al.* 1993), although not all targeting data supports this hypothesis (Deng and Capecchi 1992).

The second basic vector type is the replacement (Ω) vector, which integrates into the target site by a different mechanism to that of insertion vectors. This vector is constructed so that the linear homologous targeting sequences are interrupted by the plasmid backbone, and unlike insertion vectors, are linearised outside the region of homology. Integration of a replacement vector occurs as a result of a double reciprocal crossover event between the vector and the homologous chromosomal sequences (figure 1.5). The non-homologous plasmid sequences at the ends of the vector molecule are lost during homologous recombination events, as the vector integrates into the targeted site replacing the endogenous sequences. The major consequence of homologous recombination with a replacement type vector is loss of the endogenous homologous chromosomal sequences and their replacement with those of the targeting vector, including the plasmid backbone. This loss of chromosomal sequences means that a null allele is almost guaranteed if the

replacement event spans all or part of one or more exons, as there is no chance of generating wildtype mRNA through alternative splicing.

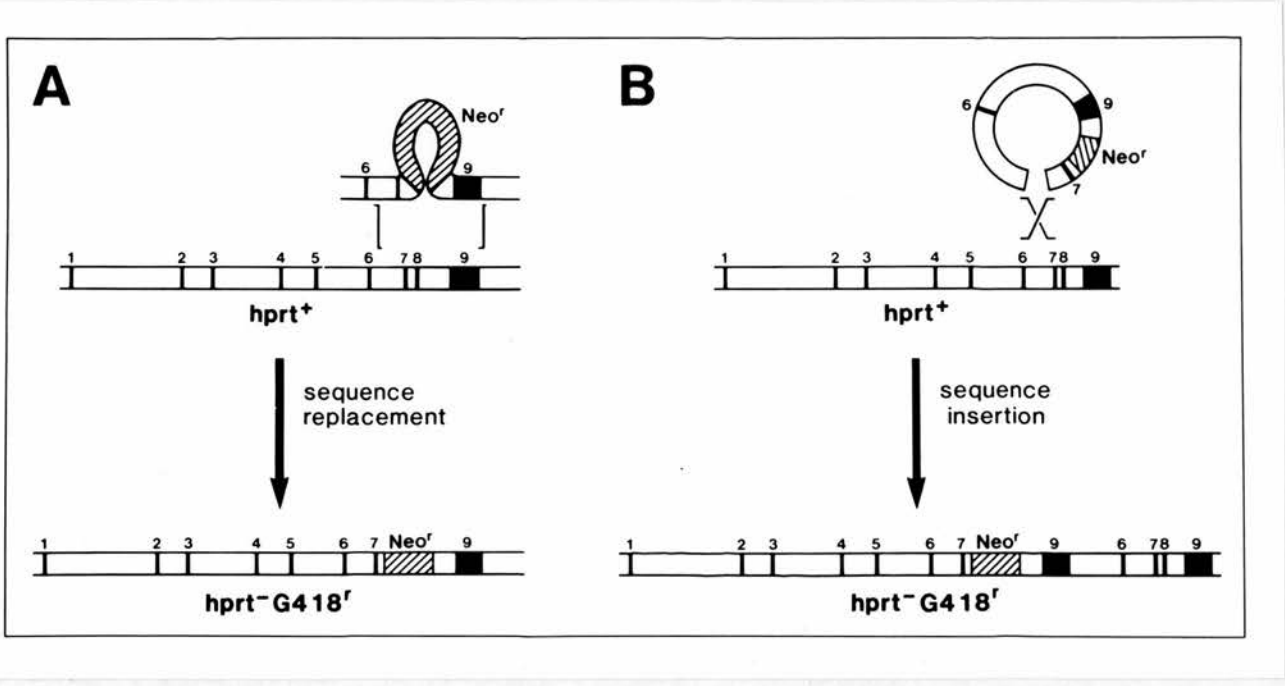


Figure 1.5 The integration patterns of insertion and replacement vectors.
Disruption of the *hprt* gene by homologous recombination with targeting vectors. (A). Replacement of *hprt* sequences on integration of a replacement (Ω) type vector through a double crossover event. (B). Insertion of an insertional (O) type targeting vector into *hprt* creating a partial duplication. The positive selection carried by the vectors is indicated by Neo^r (Neomycin), and G418 refers to the selective agent employed. Taken from Capecchi (1989).

Basic insertion and replacement targeting vectors carry a positive selectable marker which facilitates selection of transfected cells. This selection however does not discriminate between targeted (homologous recombinants) and non-targeted clones which are usually only identified after DNA analysis. By taking advantage of the different consequences of homologous and non-homologous recombination on the structure of integrated replacement vectors, Mansour *et al.* (1988) have developed a modified replacement vector to enrichment for targeted clones. This type of replacement vector carries a negative selectable gene such as *hsvtk* at one or both

ends of the vector, and is called a positive-negative selection (PNS) vector (figure 1.6). During homologous recombination, the ends of a replacement vector including the *hsvtk* sequences will be lost and are not incorporated, whereas they are retained and integrate along with the vector in a non-homologous recombination event. Therefore by applying the negative selection to the cells along with the positive selection which selects for transfected cells, it is possible to also select against randomly targeted clones which should die as they would still retain and express the negative selection gene. In practice, use of a PNS targeting vector enriches for, but does not exclusively select, homologous recombinants. The most commonly reported enrichment factor is usually around 10%.

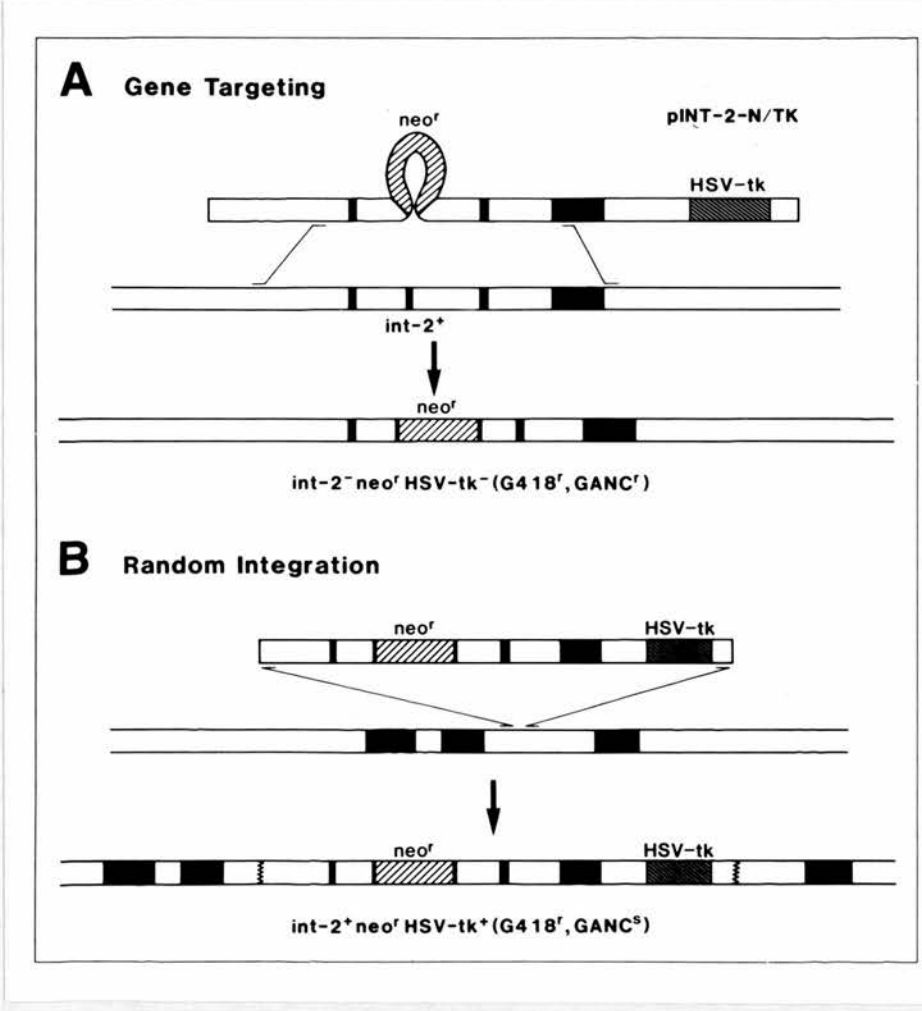


Figure 1.6 The positive-negative selection (PNS) procedure.

(A) Homologous recombination of a PNS replacement vector with the target site results in loss of the negative selection gene (*hsvtk*), as integration occurs through homologous sequences. This enables cells to survive the negative selection (Ganc). (B) Random integration occurs through the ends of the linear molecule, and results in retention of *hsvtk* sequences. Cells transfected by random integration are consequently sensitive to Gancyclovir (Ganc^s). The positive selection indicated in this diagram is Neomycin indicated by neo^r, and the selective agent is G418.

1.2.4.2 Generating subtle mutations by gene targeting.

Integration of a targeting vector into a gene through homologous recombination usually results in a major disruption of the coding sequences with the positive selection gene often acting as the mutagen. Although major gene disruptions can make a valuable contribution to our understanding of gene function, the study of subtle changes at the nucleotide level are important for a more complete functional analysis. There are a number of strategies which have been devised to introduce subtle mutations into a gene without the concurrent introduction of plasmid sequences including selectable markers. The majority of these incorporate a gene targeting step using either of the basic vector types at some stage in the process of introducing the mutation.

Mutation of the *Hox 1.1* locus is one of the few examples of gene targeting utilising a vector which did not carry a selectable marker. Zimmer and Gruss (1989) introduced a 20 base pair sequence into this locus without the inclusion of any vector sequences by introducing the DNA into the cell nucleus by microinjection and then screening for the correctly targeted cells by PCR. This is not a very common method for introducing subtle mutations as it does not include any enrichment for transfection or homologous recombination, and therefore unless this occurs at a high frequency (such as in 3% of transformants as reported by Zimmer and Gruss), it would entail screening a large number of clones to identify those bearing the desired alteration. To the author's knowledge, this experiment has not been successfully repeated at any locus.

The simultaneous transformation of cells with two vectors, one carrying a selectable marker and the other the mutant sequence, is another method of indirectly selecting for transfected cells. This method relies on the observation that some cells transformed with two unlinked DNA fragments can integrate both independently into the genome. By screening those cells which have integrated the selectable marker it is envisaged that some cells will be identified that have also incorporated the mutant DNA sequence into the target site through homologous recombination. Although such a targeted cell should also contain the selectable marker, this is likely to be

integrated at a site remote from that of the gene of interest. This method has been used to introduce mutations into the *hprt* locus (Reid *et al.* 1991., Davis *et al.* 1992). However it has the disadvantage that although non-targeted cells were frequently transformed with both vectors, correctly targeted cells rarely integrated a second DNA molecule (Reid *et al.* 1991). In addition, as the selectable marker integrates randomly into the genome, the possibility that the site of this integration might disrupt some other gene function and contribute to the phenotype cannot be excluded.

A two step recombination technique utilising insertion vectors has been successfully used to introduce subtle mutations in ES cells. This technique was first described by Scherer and Davis (1979) as a method of introducing subtle mutations into the yeast *Saccharomyces cerevisiae*, and was subsequently applied to mammalian cells by Valancius and Smithies (1991) who called it the 'in-out' targeting strategy, and Hasty *et al.* (1991) who referred to it as 'hit and run' targeting. For simplicity, it shall be referred to as 'hit and run' gene targeting from this point forward. The first stage of this technique targets the region of interest with a modified insertional vector bearing the mutation to be introduced within the homologous targeting sequences (figure 1.7). As well as the usual positive selectable marker, a 'hit and run' targeting vector incorporates a negative selection gene such as the *hsvtk* gene. Insertion of the 'hit and run' targeting vector into the target site creates a duplication of the homologous sequences separated by the plasmid sequences of the vector. Following identification of correctly targeted clones the negative selection is applied which selects against the integrated 'hit and run' targeting vector. This selection aims to select for targeted cells which have subsequently lost the targeting vector by intrachromosomal recombination through one of the duplicated regions of homology. The consequence of such a recombination event would be excision of the vector plasmid sequences along with one of the duplicated regions of homology. Depending upon the location of the crossover, vector excision could simply recreate the 'hit and run' targeting vector reverting the cell back to a wildtype genotype, or could excise the endogenous duplicated sequence along with the vector sequences, leaving the mutant sequence in its place as the only alteration to the target site. The end point of this two step

method is therefore the introduction of a subtle mutation into the target site without the presence of any plasmid sequences. The disadvantages of this method is that it requires cells to undergo two rounds of selection which increases the passage number and the opportunity for cells to lose their germ line competence. However, germline transmission was not found to be compromised when this technique was used to introduce a 14 bp insertion into the *Hoxb-4* gene (Ramirez-Solis *et al.* 1993). This targeting technique has been used successfully to introduce subtle mutations into the *hprt* locus (Hasty *et al.* 1991., Valancius and Smithies 1991), the *Hoxb-4* locus (Ramirez-Solis *et al.* 1993), and the murine type I collagen gene (Wu *et al.* 1994). These experiments reported 'run' event frequencies ranging from 3.8×10^{-3} (Hasty *et al.* 1991) to 8×10^{-7} (Valancius and Smithies 1991).

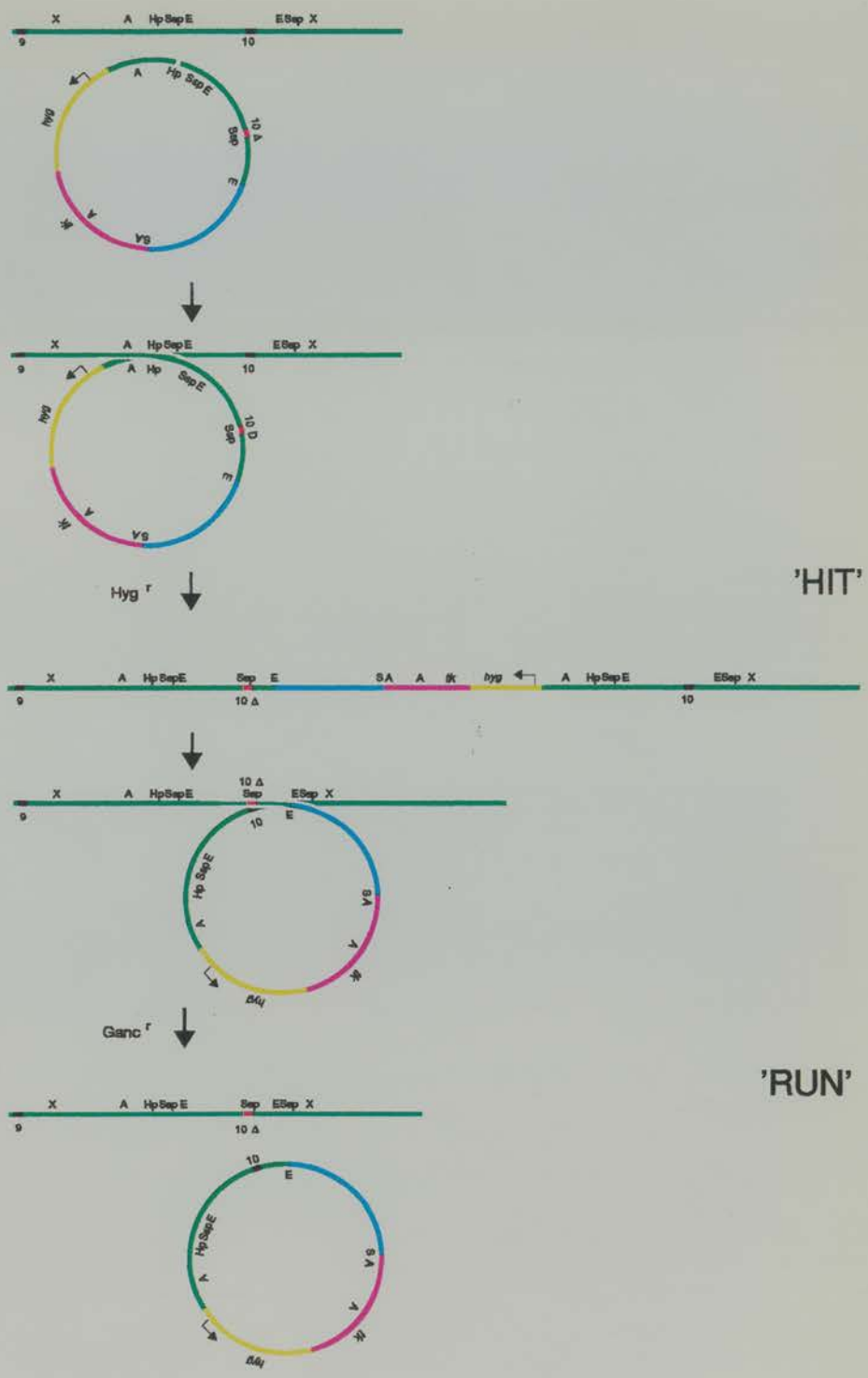


Figure 1.7 'Hit and Run' gene targeting strategy.

A two step method for introducing subtle mutations has also been devised using replacement vectors, called the double replacement targeting strategy. This strategy was proposed by Reid *et al.* (1990) and first used successfully by Stacey *et al.* (1994). This strategy involves replacement of target sequences by homologous recombination with a replacement type targeting vector. The integrated replacement vector is then targeted with sequences carrying further homology to the target site which includes the mutation to be introduced. Selection against the integrated targeting vector which carries a negative selectable marker allows cells to be obtained in which homologous recombination between the homologous sequences of the second vector and the target site has replaced the integrated vector. The net result of this being the replacement of the endogenous sequences of the target site with those bearing a subtle mutation. Askew *et al.* (1993) used a similar two step strategy called 'tag-and-exchange' targeting. With this method a vector which had inserted into the target site was replaced by mutated homologous sequences. These methods have the advantage that the same 'marked' allele created by the first targeting step can be replaced by many different mutant homologous sequences in subsequent targeting experiments and so need only to be targeted once. However, it is not advisable to use just one targeted line to generate all transgenic animals as there is always the possibility that this might have an additional mutation which could affect the phenotype of the resulting mice.

Other novel ways of engineering subtle mutations have been devised. Deng *et al.* (1993) used a replacement type vector to replace exon 10 of the murine *Cftr* gene with a mutant exon bearing the ΔF508 mutation, with the vector sequences being present in the non-coding intron. This method has the disadvantage that the final structure of the targeted site includes plasmid sequences and an expressing selectable marker which may interfere with expression of the endogenous gene. Use of the Cre/*loxP* recombinase system has also been proposed as a means of removing vector sequences following targeting of mutations to regions of interest (Baringa 1994). With this system, the region of interest would be targeted with a replacement vector bearing the mutation to be introduced. The presence of *lox P* flanking the plasmid

sequences would lead to its subsequent deletion on expression of the Cre enzyme within the cell, leaving only a 34 bp sequence in the intron.

It is also possible to combine classical transgenic techniques with gene targeting to study a mutant protein. Transgenic mice produced by microinjection of non-targeted DNA can express a mutant form of a gene at a site remote from the endogenous gene. By crossing these mice with those which have had the endogenous alleles inactivated by gene targeting, it is possible to get the mutant transgene onto a null background. In these animals, the mutant protein of the transgene will be present as the only product from this gene, and although not directly analogous to a subtle mutation, should give some valuable information on the nature of the mutant gene product.

1.2.5 Animal Models For Human Genetic Disease Produced By Gene Targeting

The first human genetic disease to be modelled in the mouse through use of ES cells was Lesch-Nyhan syndrome. This inherited disease is caused by a deficiency of the enzyme hypoxanthine phosphoribosyltransferase (HPRT) and is characterised by mental retardation and behavioural abnormalities such as self-mutilation (Lesch and Nyhan 1964). Utilisation of a selection against *hprt*, enabled ES cells lacking a functional *hprt* allele to be obtained which had arisen spontaneously (Hooper *et al.* 1987), or as a consequence of retroviral insertion from random mutagenesis (Kuehn *et al.* 1987). These ES cells were used to generate mice lacking any HPRT activity which it was anticipated might closely model the human disease. However, although mice homozygous for the null allele did not exhibit any HPRT activity, they displayed none of the defects associated with the human condition and were as far as could be determined, phenotypically normal. It appeared that the mouse is less dependant upon HPRT for purine salvage than humans, and predominately uses another enzyme, adenine phosphoribosyltransferase (APRT) for this process. Therefore an absence of the enzyme HPRT has much more severe consequences for humans than mice. This was borne out by a later study which showed that these HPRT deficient mice could be induced to exhibit some of the hallmarks of Lesch-Nyhan syndrome by inhibition of APRT activity (Wu and Melton 1993). Following administration of APRT inhibitors,

the mutant mice displayed self-injurious behaviour, principally through over-grooming, reminiscent of that displayed by Lesch-Nyhan patients.

The first reported animal model for a human genetic disease created by gene targeting was for Gaucher's disease (Tybulewicz *et al.* 1992). A targeted disruption of the murine glucocerebrosidase gene generated mice deficient in the enzyme responsible for sphingolipid glucocerebroside degradation, the cause of Gaucher's disease in humans. This produced a severe phenotype in homozygous mice which died within twenty four hours of birth, and is analogous to the more severely affected, type two acute neuronopathic subset of Gaucher's patients who also die shortly after birth.

Mutation and allele loss of the p53 tumour-supressor gene has been associated with tumours in a wide range of human organs and is also associated with an inherited increased cancer-susceptibility syndrome called Li-Fraumeni syndrome. The p53 protein is implicated in cell cycle control and was therefore thought to be essential for embryonic development. However when the gene was inactivated by gene targeting in ES cells, it produced viable, phenotypically normal homozygotes (Donehower *et al.* 1992). This suggests that in these mice, the function predicted to be carried out by p53 was accomplished by other proteins. This may indicate that p53 does not have a role in cell cycle control during development, or that other proteins or pathways were able to compensate for its deficiency in these mice. Although developmentally normal, these mice exhibit a higher than normal susceptibility to spontaneous tumours and this phenotype may be analogous to that of the Li-Fraumeni syndrome.

The phenotype resulting from a targeted gene alteration cannot always be predicted and many different factors may influence the physiological consequences. As with p53 and HPRT, cells might possess other pathways or mechanisms which could compensate for the loss of activity resulting from the introduced mutation. This compensatory effect might also come from other sources. The transforming growth factor TGF-β1 was thought to have a vital role during development and therefore a null allele was predicted to be an embryonic lethal mutation in homozygotes.

However, mice homozygous for a mutation introduced by gene targeting were indistinguishable from wildtype littermates at birth, but developed histological abnormalities 7 days later (Shull *et al.* 1993). Subsequent studies have shown that this phenotype is a result of a maternal source of TGF-β1 'rescuing' the phenotype of the targeted gene disruption up until birth (Letterio *et al.* 1994). Yagi *et al.* (1994) found a variation in the survival of homozygotes for a null allele created in the *fyn* tyrosine kinase gene, which depended upon the genotype of the parents. The homozygous offspring of heterozygous parents had no phenotype, whereas the homozygous pups of homozygous parents died through an inability to suckle. This was found to be due to the inability of the homozygous pups to stimulate lactation in the female, which was not a problem in mixed genotype litters, as littermates with a wildtype allele could activate lactation in the mother, thus enabling the homozygotes to survive. If the targeted gene has an essential phenotype during embryonic development, a high ES cell contribution to organs and tissues of a chimaera might induce an embryonic lethal phenotype. This was seen in chimaeras produced by microinjection of ES cells containing a null mutation in the X-linked, methylated DNA binding protein MeCP2 (Tate *et al.* 1994). The higher the ES cell contribution, as judged by expression of a reporter gene carried by the targeting construct, the more abnormal the embryo development *in utero*. The results of these and other targeting experiments have therefore demonstrated that it is often not possible to predict the phenotype resulting from a targeted gene alteration.

1.2.5.1 Mouse models for cystic fibrosis

There have been four models for cystic fibrosis created by gene targeting in ES cells to date. All four of these targeting strategies sought to disrupt *Cftr* function through integration of a targeting vector.

Koller *et al.* (1991) created a targeted disruption of *Cftr* by targeting the region around exon 10 with a PNS replacement vector. This targeting vector replaced part of exon 10 with selectable markers, and incorporated a stop mutation in the 3' homologous targeting sequences. Mice homozygous for the disrupted gene exhibited

a severe phenotype with a large number dying before or around weaning (Snouwaert *et al.* 1992). A high percentage of these perinatal deaths were caused by intestinal obstruction and blockage, similar to the meconium ileus condition seen in humans. However, meconium ileus only occurs in approximately 10% of CF cases and therefore this condition is exacerbated in the mouse. These mice exhibited many of the features of CF such as the alteration of mucous and serous glands, and obstruction of these glands with thickened mucus. Electrophysiological assays demonstrated a defect in chloride ion transport which was analogous to the defect seen in humans (Clarke *et al.* 1992). Both male and female mice homozygous for the null mutation were fertile in contrast to the human condition, where females have reduced fertility and only 5% of males are fertile.

A similar null mutation was created by Ratcliff *et al.* (1992). They also replaced some of the sequences around exon 10 using a PNS replacement vector. Mice homozygous for this mutation also exhibited the severe intestinal disease exhibited by the 'knockout' mice of Snouwaert *et al.* (1992), (Colledge *et al.* 1992., Ratcliff *et al.* 1993), with approximately 80% dying within one week of birth. The phenotype of these mice was very similar overall to that of the other CF mice created by targeting with a replacement vector, with mucus accumulation reported in many glands

Another disruption of *Cftr* was created, this time using an insertion vector to target the region around exon 3 (O'Neal *et al.* 1993). Insertion of this vector created a duplication of exon 3 separated by plasmid sequences, which produced a similar severe phenotype in homozygotes to that created by the mutation of exon 10 by replacement vectors. As with the other mouse models, a high rate of perinatal death was observed with 40% dying within the first week from intestinal obstruction. Mucus accumulation was again observed in the crypts of the small intestine and in many glands.

All of the above CF mouse models exhibited a similar phenotype of severe intestinal disease which was more pronounced than in the human condition.

Electrophysiological assays confirmed the chloride channel defect characteristic of CF, however very little if any disease was noted in the main sites of disease in man, namely the lungs and the pancreas. Clarke *et al.* (1994) found a correlation between the presence of an alternative, calcium-regulated chloride channel and the absence of disease in the 'knockout' mice of Snouwaert *et al.* In organs severely affected in these mice, such as the intestine, no chloride transport by this alternative channel could be detected, whereas it was present in the mildly affected pancreas and pulmonary epithelia. This suggests that in some tissues of the mouse, this calcium-regulated channel can compensate for absence of CFTR resulting in a lack of pathology in these organs of the null mice.

A different mutation was created by Dorin *et al.* (1992) using an insertion vector to target exon 10. Integration of this targeting vector created a partial duplication of exon 10 fused in frame to a gene for neomycin resistance. The presence of the poly A tail of the neomycin gene and the occurrence of stop mutations in all three reading frames in the plasmid sequences were predicted to cause premature termination of transcription, causing a null mutation. However the phenotype of mice homozygous for this mutation was milder than that described by other groups, with a reduced incidence of meconium ileus (10%), more in line with that reported for the disease in humans (Dorin *et al.* 1992). These mice also exhibited the defective chloride ion transport of CF and could be distinguished from their normal littermates solely on this basis. Again there was an absence of pancreatic disease, and a detailed study of the chloride ion conductance detected significant chloride transport by the calcium-regulated chloride channel in the pancreas of these mice (Gray *et al.* 1994). These mice have improved long term survival, and in contrast to those with a severe phenotype, can be distinguished in tracheal Ussing chamber studies from their wildtype littermates. In addition, lung disease can be precipitated following bacterial challenge (Dorin *et al.* 1994). The milder intestinal phenotype of these mice is due to the residual wildtype mRNA present in these mice (approximately 10%), and is perhaps analogous to one of the milder class of human CF-causing mutations. Mutant CFTR bearing these types of mutations retains some degree of function, typically 15-

30% of wildtype (Sheppard *et al.* 1993), and manifests as severe lung disease but with pancreatic sufficiency. In addition, CF patients with a similar 'leaky' mutation have recently been described who also display approximately 8% residual wildtype *CFTR* mRNA, and exhibit severe lung disease (Highsmith *et al.* 1994). A similar phenotype amelioration has been described by Moens *et al.* (1992). They also created a 'leaky' insertional mutation to ameliorate the embryonic lethal phenotype of a null mutation in the N-*Myc* gene (Sawai *et al.* 1991). Homozygotes for the leaky insertion mutation survived to birth, in contrast to the death at mid-gestation of the null homozygotes.

These mouse models for CF are a valuable resource for CF research. The similarities and differences in the phenotypes when compared to the human condition should yield important information regarding the aetiology of the disease. The identification of the apparent compensatory effect of the calcium-regulated chloride channel in some tissues in these mice may have important consequences for future CF therapy. By studying the electrophysiology of these mice which lack any CFTR, a deficit in function of another chloride channel has been identified indicating a link between the two (Gabriel *et al.* 1993). The hypothesis of an increased resistance to the cholera toxin in CF heterozygotes to account for an 'heterozygote advantage' effect has been given credence by testing the effects of this toxin in CF mice (Gabriel *et al.* 1994). One of the most valuable applications of the CF mouse models will be for testing new therapies. Already studies have been reported in which potential gene therapies for humans were evaluated in the mice (Hyde *et al.* 1993., Alton *et al.* 1993., Grubb *et al.* 1994). The opportunity to evaluate therapies in a CF mouse is of great benefit, and approval for phase I trial in humans were subsequently obtained based upon the promising results of Alton *et al.* The CF mouse models have therefore already been of great benefit to the CF community.

1.3 PROJECT AIM

The aim of this project was to create mouse models for cystic fibrosis which bear precise CF-associated mutations as the only alteration to the murine *Cftr* gene.

Cystic fibrosis has already been modelled in the mouse by disruption of *Cftr* with vector sequences which produce absolute or 'leaky' null mutations. These animals have displayed many hallmarks of the disease and are a valuable resource for CF research. The majority of CF-causing mutations in humans however, do not create null alleles. Over 400 CF-associated mutations have been described to date, and the predominant mutation (ΔF508), present on approximately 70% of CF chromosomes, is a three base pair deletion. The severity of CF varies enormously between affected individuals and a correlation with genotype has only been found for pancreatic involvement, and not for the lung disease which is the major cause of premature death. Therefore the study of 'knockout' mutant mice cannot resolve some of the important issues surrounding CF, such as the effect of different mutations on *CFTR* function and the genotype phenotype relationship. The availability of mice which bear precise, clinically relevant mutations and therefore more closely model the human disease, might allow some of these important issues to be addressed. The inbred background and the controlled environment of the laboratory mouse would allow the contribution of these factors to the disease phenotype to be ascertained. It would also be of interest to determine if the presence of a mutant protein had a different phenotypic consequence to the complete absence as seen in the null mice. In addition, these mice could be used for testing potential therapies. Those which posses mutations causing the partially functional mutant protein to become mislocalised (including the most common ΔF508 mutation) would be invaluable for testing relocation strategies.

I aim to introduce CF-associated mutations into the murine *Cftr* gene by using the 'hit and run' gene targeting technique in ES cells. This targeting technique was chosen as it typically introduces a precise mutation into the target site without any accompanying vector sequences. Furthermore it utilises insertion type targeting

vectors which have already been used in our lab to achieve high frequency targeting of the murine *Cftr* gene in ES cells. Mice generated by these techniques would then be used to study many aspects of the disease, and especially to test strategies for relocating mutant proteins subject to biosynthetic arrest. These mice should be a valuable asset to CF research.

CHAPTER 2

MAMMALIAN CELL CULTURE MATERIALS AND METHODS

2.1 EMBRYONAL STEM CELL CULTURE

2.1.1 Cell Lines

The embryonal stem (ES) cells used in this study were either E14 ES cells (Handyside *et al.* 1989), a kind gift from Dr Martin Hooper, or CGR8 ES cells, a kind gift from Dr Austin Smith. Both of these cell lines have an XY karyotype, were derived from the mouse strain 129/Ola and carry the following marker alleles:

Agouti (A) Locus: wild type (A^W)
Albino (c) Locus: chinchilla (c^{ch})
Pink Eye (P) Locus: pink-eyed dilution (p)

2.1.2 Culture Media And Solutions

General considerations

In the choice of culture media and reagents, a number of precautions were taken to avoid the exposure of ES cells to agents which may induce differentiation. Glassware detergent constitutes one such agent and therefore all medium and solutions were either obtained from commercial sources, or made up in detergent-free containers. Disposable, sterile pipettes (Costar) were used throughout. The quality of the water used in solutions is also an important consideration. All solutions prepared in-house utilised ultrapure water from an Elgastat UHP reverse osmosis filtration system. Solutions were sterilised either by autoclaving at 121°C for 20 minutes, or by filtration through a 0.2µm sterile filter (Gelman). Prior to use, all solutions were pre-warmed to 37°C.

Medium

Both the E14 and the CGR8 cell lines were routinely maintained in Glasgow's modification of Dulbecco's Medium (GMEM) obtained as a single strength solution, initially from Flow-ICN laboratories and later from Gibco-BRL. This medium was supplemented as follows:

15% batch-tested foetal calf serum (FCS, Initially supplied by Globepharm Ltd and subsequently by Gibco-BRL)

0.3% glutamine (Gibco-BRL)

1 mM sodium pyruvate (Sigma)

0.1 mM MEM non-essential amino acids(Sigma)

0.1 mM β-mercaptoethanol (Sigma)

The above medium was also supplemented with Leukaemia Inhibitory Factor (LIF) where the ES cells were to be maintained in the absence of feeder layers (section 2.13). As a rule, medium was supplemented at least 24 hours prior to use, and its sterility checked by incubating 1 ml overnight at 37°C in 10 ml Tryptose Phosphate Broth (Oxoid).

Phosphate buffered saline (PBS)

PBS was made up by dissolving one calcium and magnesium-free PBS tablet (Gibco-BRL) in 100 ml of ultrapure water. Sterilisation was by autoclaving.

Trypsin-EDTA solution

This was obtained as a sterile solution at ten times normal strength, consisting of 0.5% trypsin and 0.2% EDTA in 0.9% Sodium chloride (Sigma). This was routinely stored at -20°C, diluted to normal strength in sterile PBS when required, and stored at 4°C for up to two weeks.

Gelatin solution

A sterile solution of 0.1 % gelatin was used to coat surfaces prior to the culture of ES cells in the presence or absence of feeder layers. This was prepared by adding the appropriate amount of tissue culture grade, porcine skin gelatin (Sigma) to ultrapure water and autoclaving. The sterile solution was stored at 4°C.

2.1.3 Production Of Leukaemia Inhibitory Factor (LIF)

ES cells can be maintained in an undifferentiated state *in vitro* in the absence of feeder layers if LIF (also known as Differentiation Inhibitory Activity-DIA) is added to the culture medium (Smith *et al.* 1988). Recombinant LIF was produced by overexpression of LIF cDNA in the eukaryotic expression vector pC10-6R, (a kind gift from Dr Austin Smith), in Cos-7 monkey kidney cells. The Lipofection method of transfection was used, as in our hands it routinely gives high frequency and often multicopy transfection. Lipofectin (Gibco-BRL) was the lipofection agent employed, and consists of a 1:1 (w/w) mixture of the cationic lipid (N[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride) (DOTMA) and dioleoyl phosphotidylethanolamine (DOPE). This reagent forms cationic liposomes, and can interact with the negatively charged DNA to form liposome-DNA complexes capable of delivering DNA into a cell upon fusion with the cell membrane.

2.1.3.1 Transfection of Cos-7 cells

1. Cos-7 cells were grown to 70-80% confluence in 225 cm^2 vented-cap tissue culture flasks (Costar), as described in section 2.2.3.
2. The cells were washed twice with serum-free DMEM and covered with 40 ml of serum-free DMEM for incubation.
3. For each flask, 120 µg pC10-6R plasmid was diluted to a total volume of 300µl with sterile distilled water, and gently mixed with 440 µl Lipofectin (Gibco-BRL).
4. The DNA-liposome mixture was left to incubate at room temperature for 15 minutes to allow the formation of DNA-Lipofectin complexes, after which time it was added slowly, dropwise to the culture medium.
5. Following a 6 hour period of incubation at 37°C, 40 ml of DMEM containing 20% FCS was added to the flask, reducing the overall FCS concentration to 10%.
6. After further incubation at 37°C for 48 hours, the LIF-containing medium was harvested, centrifuged at 1000 rpm (Sorvall Coolspin centrifuge) to remove any debris, and filter-sterilised.

7. The LIF was aliquoted and stored at -70°C.

2.1.3.2 Titration of LIF activity

In order to determine the lowest concentration at which LIF effectively inhibited the differentiation of ES cells, a dilution series was set up and evaluated as follows:

1. Twenty pre-gelatinised 25 cm^2 tissue culture flasks were seeded with $5x10^5$ cells per flask.
2. The following dilutions of LIF were added to flasks in duplicate, 0, 1/100, 1/250, 1/500, 1/750, 1/1000, 1/2000, 1/5000, 1/10 000. LIF of a known activity was added at an effective concentration to two flasks as a positive control.
3. Cells were cultured for two weeks and then assessed for differentiation by fixing with methanol and staining with Giemsa (BDH) diluted 1 in 10 in Gurr's buffer (BDH).
4. The highest dilution which supported the undifferentiated growth of ES cells was used as the working concentration. This concentration was commonly 1/500-1/1000 dilution.

2.1.4 Preparation of Feeder Layers

ES cells were maintained, whenever possible, on feeder layers of mitotically inactivated mouse primary embryonic fibroblasts (PEFS). Both E14 and CGR8 ES cell cultures were found to display a lower degree of differentiation when grown on feeder layers. The methods of feeder layer preparation employed were adapted from Robertson (1987) and were as follows:

1. A female Swiss albino mouse was sacrificed at thirteen to fourteen days of pregnancy.
2. The abdomen was swabbed with 70% ethanol, and the uterine horns dissected out using sterile instruments and aseptic technique.
3. The uterus was placed into cold PBS (approximately 4°C) on ice for transfer to tissue culture facilities.

4. The uterine horns were placed in a 100 mm petri dish in 10 ml of prewarmed PBS and the embryos dissected away form the uterus and the foetal membranes, and quickley killed by decapitation.
5. The embryos were transferred to a fresh plate containing 10 ml of PBS and soft tissues such as the heart and liver were dissected out from the embryonic tissue using sterile, fine watchmakers forceps.
6. Each embryo was then placed into an individual well of a 24 well plate containing 1 ml of Trypsin-EDTA, and finely minced using sterile, fine iris scissors.
7. This embryonic tissue was incubated at 37°C for 5 minutes to facilitate its enzymatic disaggregation by trypsin.
8. When this had occurred, the trypsin was neutralised by addition of 1 ml of DMEM. The supernatant was collected from all of the wells, pooled, and made up to 50 ml with DMEM.
9. After standing for 5 minutes to allow large pieces of cellular debris to settle out, the cell suspension was seeded into a 175 cm^2 tissue culture flask and incubated at 37°C in 5% CO_2 for 24 hours.
10. The following day, the medium was changed to remove cellular debris, and the fibroblasts incubated further until confluent.
11. Once confluent, the cells were split between twenty 175 cm^2 tissue culture flasks and incubated again until confluency was achieved.

At this stage the fibroblasts were harvested and mitotically inactivated by one of two methods:

2.1.4.1 Mitotic Inactivation Of PEFS By Gamma Irradiation.

This was the preferred method of mitotic inactivation as it was inexpensive, and there is no danger of carryover of mitotic inhibitor as with the Mitomycin C method.

A single cell suspension of fibroblasts in 4 ml of DMEM was placed into a sterile bijou bottle and lowered over a ^{60}Co gamma radiation source. The cells were exposed

to a dose of 3000 rads determined by length of exposure to the source (approximately 55 minutes), and then prepared for freezing at -180°C by the usual method (section 2.1.11).

2.1.4.2 Mitotic inactivation by Mitomycin C.

The medium was aspirated from confluent flasks of fibroblasts and replaced with 40 ml DMEM containing 10 µg/ml of mitomycin C (Sigma). The cells were incubated for 2 to 3 hours, the medium was then aspirated, and the cell monolayer washed three times with PBS. The fibroblasts were harvested and prepared for freezing at -180 °C by the usual method (section 2.1.11.1).

Each batch of PEFS was tested for mitotic inactivation by plating out a small number of cells at a sub-confluent density and growing these on for seven days to ensure confluency was not achieved. The inactivated fibroblasts were stored in a liquid nitrogen freezer at -180°C until required. When required, the PEFS were recovered from liquid nitrogen by the standard procedure (section 2.1.11.2), seeded into a gelatinised flask and cultured in GMEM. It was not strictly necessary to establish a PEF monolayer prior to seeding ES cells, and the PEFS could be seeded at the same time or even at a later date.

2.1.5 Maintenance of ES Cells.

2.1.5.1 General considerations

ES cells were routinely maintained in the absence of antibiotics and therefore the sterile integrity of the equipment and reagents used, as well as the sterile technique of the operator, were of paramount importance. To this end, all manipulations of ES cells were carried out in a Biomat class II laminar flow cabinet which was swabbed with 70% ethanol prior to, and after use. Anything to be introduced into this sterile environment was also swabbed with ethanol. To minimise the risk of contamination from the operator, a dedicated lab coat was always worn as well as latex gloves which were swabbed frequently with ethanol.

Tissue culture flasks for culturing ES cells were obtained from Costar and had a 0.2 μ m filter incorporated into the cap for extra sterility. All multiwell plates were obtained from Falcon, and 100 mm plates from Costar.

To maintain the desired pluripotential phenotype, ES cells were grown at relatively high densities and fed daily (or when the acidity of the medium dictated). ES cells grew rapidly, dividing every 18-24 hours, and were routinely sub-cultured at 3-4 day intervals.

2.1.5.2 Sub-culture of ES cells.

ES cells were passaged as soon as confluency was achieved to avoid medium exhaustion and cell death.

1. A confluent 25 cm^2 flask was typically passaged by aspirating the medium and washing the cells with 5 ml of PBS.
2. After removal of the PBS, 0.5 ml of Trypsin-EDTA was added and the flask incubated at 37°C for 5 minutes.
3. Cells were dislodged from the flask bottom by gently tapping the flask and were examined under low magnification on an inverted light microscope to ensure that a single cell suspension had been achieved, (cellular aggregates are undesirable as these can differentiate to form endoderm-like cells).
4. The trypsin was inactivated by addition of 9 ml of GMEM, and the cell suspension spun down in a sterile universal tube (Sterilin) at 1000 rpm (Sorvall Coolspin).
5. The supernatant was aspirated, and the cells resuspended in 10 ml of GMEM.
6. Cell density was determined using a haemocytometer, and cells were plated into a pre-gelatinised 25 cm^2 flask (plus or minus PEFS) at a density of 1x 10^6 cells.
7. These were grown in an incubator at 37°C in an atmosphere of 5% carbon dioxide.

For larger or smaller scale cultures the volumes were scaled up or down proportionately.

2.1.6 Karyotype Analysis of ES Cells

It is important to routinely check the karyotype of ES cells as cell populations can drift and aneuploid cells can become selected over periods of continuous culture. ES cells were routinely checked for the correct chromosome number (40) by the preparation and staining of mitotic spreads. In the later stages of this project this was carried out by Fiona Kilanowski. The method employed is essentially that of Robertson (1987) and was as follows:

1. An exponentially dividing culture of ES cells in a 25 cm^2 flask was mitotically inactivated by adding 0.02 μg Colcemid (Sigma) to the medium and incubating for one hour.
2. The cells were harvested and spun down by the usual procedure (section 2.1.5).
3. After aspiration of the medium, the cells were resuspended initially in 1 ml of 0.56% Potassium chloride (KCl), and then 6 ml of KCl.
4. The cells were left to stand for 10 minutes at room temperature, and then spun down at 500 rpm (Sorvall Coolspin).
5. KCl was removed and replaced with ice-cold fixative consisting of 3 volumes of methanol to 1 volume of glacial acetic acid.
6. After standing at room temperature for a further 5 minutes, the fixative was changed another 3 times and the cells were resuspended in a final volume of 1 ml.
7. Chromosome spreads were prepared by dropping one drop of the cell suspension from a Pasteur pipette (at height of approximately 10 cm) onto a microscope slide which had been pre-cleaned in ethanol.
8. Slides were air-dried and stained by immersion for 15 minutes in 3% Giemsa stain (BDH) diluted in PBS.
9. Two rinses in distilled water were used to remove the excess stain, and the slides air dried before examination under a microscope.

2.1.7 Mycoplasma Detection

It is important to ensure that ES cells are free from mycoplasma contamination as infected cells do not grow well and contribute poorly to the germ line of chimaeras (Bradley 1987). Mycoplasma contamination is commonly transmitted by poor aseptic technique or the use of contaminated reagents, and is often concurrent with bacterial or fungal infection. To facilitate early detection of contamination, all ES cell culture was carried out in the absence of antibiotics which can mask contamination by keeping it at a low level. All reagents were bought from commercially tested sources, and LIF made 'in house' was tested prior to use.

Each ES cell line was tested regularly for mycoplasma contamination by one of the following methods:

(i) Hoechst staining

This method utilises Hoechst stain, a fluorescent stain which binds specifically to DNA, to indicate the presence of mycoplasma infection of ES cells. Mycoplasma testing was carried out by Mr Bill Christie using the method of Chen (1977). This involved the culture of cells on a glass microscope slide which were then fixed, stained, and examined using fluorescence microscopy. Mycoplasma infection was indicated by the presence of fluorescent bodies in the cytoplasm and nucleus of the cell.

(ii) Direct culture

The presence of mycoplasma was also detected by the presence of characteristic mycoplasma colonies when contaminated cell suspensions were plated out onto media which is selective for mycoplasma. This method of testing was also carried out by Mr Bill Christie using the method described by Barile (1973).

(iii) PCR analysis

In the later stages of this work, all mycoplasma testing was carried out by Fiona Kilanowski using Stratagene's Mycoplasma PCR primer kit. This utilised

mycoplasma-specific primers to amplify mycoplasma DNA during a PCR reaction on cell extracts. It included positive and internal controls and provided a very quick, simple and sensitive method of mycoplasma detection.

1. Cultures to be tested were grown to near confluence and 100 µl of medium removed into a microfuge tube for testing. This was placed in a heat block at 95°C for 5 minutes
2. The boiled cell extract was mixed with 10 µl of StrataClean resin (Stratagene) and spun briefly to pellet the resin.
3. An aliquot of the supernatant was removed, diluted 1:10 in UV-irradiated sterile distilled water, and 10 µl used for the PCR reaction.
4. A PCR reaction mixture was made up in 40 µl with 5 µl 10x Taq buffer (Perkin Elmer), 2 mM Magnesium chloride (Perkin Elmer), 200 µM each dNTP (Pharmacia) and 1 unit AmpliTaq polymerase (Perkin Elmer). Reactions were set up on boiled cell extracts and incorporated a negative control as well as the appropriate positive and internal controls (provided with the primer kit).
5. The PCR amplification was carried out on a Hybaid Omnigene PCR thermal cycler using the following program:

SEGMENT	CYCLES	TEMPERATURE	TIME
1	1	94 °C	5 minutes
		55 °C	1.45 minutes
2	3	72 °C	3 minutes
		94 °C	45 seconds
		55 °C	1.45 minutes
3	40	72 °C	3 minutes
		94 °C	45 seconds
		55 °C	45 seconds
4	1	72 °C	10 minutes
		27 °C	10 minutes

6. The PCR products were resolved by electrophoresis through a 3% agarose gel as described in section 3.4. Mycoplasma contamination was indicated by the presence of appropriately sized bands.

2.1.8 Electroporation of ES Cells

Transfection of ES cells with DNA was achieved by use of the electroporation technique (Chu *et al.* 1987). Many transfection methods are available, but for these experiments electroporation was the method of choice as it is quick, simple, consistent, and in our hands usually transformed cells with a single copy of DNA.

A high voltage electrical pulse is applied to cells which allows DNA to enter via pores in the cell membrane. The conditions for optimal transfection efficiency typically result in 50% cell death. The parameters influencing transfection efficiency and cell death are voltage, ionic concentration, DNA concentration and cell density. The

transformation efficiency is low and to compensate for this the transforming DNA carries a selectable marker gene.

The conditions employed in this work were as follows:

1. Recently thawed ES cells were grown up (usually on PEFS) until 80% confluency in a 75 cm^2 tissue culture flask was achieved.
2. Cells were trypsinised and counted in the usual fashion, spun down and resuspended in PBS to give a cell density of 1.2 x 10^7 cells/ml.
3. A volume of 0.8 ml of the cell suspension was mixed with 100 µl of linearised vector, transferred to an electroporation cuvette (BioRad) and left to stand for 10 minutes.
4. The cuvette was placed into the cuvette holder and inserted into the electroporation chamber of the electroporation machine (BioRad Genepulser). The settings were adjusted to 800 V, 500 µF capacitance, and the electric pulse delivered. If operating correctly, a time constant of 0.1 - 0.2 was observed.
5. Cells were left to stand for 10 minutes, diluted into 10 ml of GMEM and seeded into ten 100 mm gelatinised dishes containing 20 ml GMEM. The plates were incubated as usual.
6. The medium was changed and the appropriate drug selection imposed 24 hours after electroporation.
7. Every second day the medium and selection drug were changed until colonies were visible and ready for picking.

2.1.9 Drug Selection

2.1.9.1 Positive selection markers.

As mentioned previously, electroporation was the chosen method of transfection for ES cells but gives a relatively low transfection efficiency. Therefore, positive selection markers were used to isolate the minority of transfected cells from the majority of non-transfected cells.

The positive selection markers used in this work were Neomycin phosphotransferase which was selected for using the aminoglycoside G418 (Gibco, also known as Geneticin), and Hygromycin B phosphotransferase which was selected for using the aminoglycoside Hygromycin B (Calbiochem). Both are dominant selectable markers of bacterial origin and encode resistance to these aminoglycosides through their inactivation by phosphorylation. In non-resistant cells, G418 binds to the 16s rRNA and Hygromycin B binds to the 30s rRNA, interfering with and inhibiting translation and hence protein synthesis (Davies *et al.* 1993).

2.1.9.2 Negative selection markers.

Negative selection markers were used to eliminate transfected cells expressing a negative selection gene carried by a vector, from a mixed population of expressing and non-expressing cells. Throughout these experiments, the negative selection marker employed was the Herpes Simplex virus Thymidine Kinase (*hsvtk*) gene. This is a dominant selectable marker and was selected against using Gancyclovir (Syntex) but 1(1-2-deoxy-2-fluoro-β-D-arabinofuransyl)-5-iodouracil (FIAU) can also be used. In both cases, a toxic base analogue is metabolised by *hsvtk* (but not the endogenous mammalian *tk* gene), which has fatal consequences for the cell. Due to the toxic nature of the *hsvtk* gene product, neighbouring, non-expressing cells can be also killed by the 'bystander effect' when the toxic product is passed from expressing cells. This effect was minimised by plating the cells at a low density (usually 2 x 10^3 cells/cm^2) for Gancyclovir (Ganc) selection.

2.1.9.3 Determination of effective drug selection concentration.

It is important to determine the effective drug concentration for each individual cell line so that the minimum effective dose can be used, and its influence on viability and pluripotency of the cells can be checked.

The minimum effective dose was determined as follows:

1. ES cells were plated into 10 wells of a 24 well plate at a density of 1 x 10^4/well.

2. Twenty four hours after plating, a dilution series based upon reported working concentrations of the selection drug was added to the wells. The cells were incubated as normal.
3. The medium and the selection drug were replaced every two days, and the cells cultured until differences between dilutions were observed (at least one week).

For positive selection, the working concentration was taken as the lowest concentration giving death of 100% of wild-type cells. For negative selection, the working concentration was taken as the lowest concentration giving death of 100% of cells transfected (and expressing) the *hsvtk* gene, which had no deleterious effect upon wild-type cells.

For the cell lines employed in these experiments, these concentrations were as follows:

- 200 µg/ml G418 *
- 150 µg/ml Hygromycin B
- 2.5 µM Gancyclovir

* It was necessary to test each new batch of G418 as there can be considerable batch variation.

2.1.10 Clone Picking And Expansion

Once clones were obtained, they had to be picked and maintained as separate cell lines. Each clone was picked, and grown up to generate cells for DNA analysis and also for frozen stocks which could be recovered upon identification of clones with the desired genotype. This was carried out as follows:

1. Medium was aspirated from a plate containing clones which was then washed with 10 ml of PBS. The PBS was aspirated and replaced with a further 5 ml of PBS.
2. Colonies to be picked were encircled on the underside of the dish using a fine tipped marker pen and the total number of clones per plate noted.

3. Clones were observed under low power using an inverted microscope placed inside a laminar flow cabinet.
4. Each clone was picked off using a sterile, fine-tipped pastette (Alpha Laboratories Ltd) and placed into an individual well of a 96 well bacteriological plate containing approximately 150 µl of Trypsin-EDTA. This was incubated for 5 minutes at room temperature.
5. After incubation, the clone was pipetted up and down to facilitate trypsinisation, and incubated again for a further 5 minutes.
6. Once single cell suspensions had been achieved, the clones were seeded into 1-2 wells of a 96 well plate and grown up without selection (in most cases on a feeder layer of PEFS).
7. When confluent, clones were trypsinised and seeded into fresh wells for expansion.
8. At the earliest opportunity, at least one confluent well of a 96 well plate was frozen away at -180 °C in liquid nitrogen storage.
9. Cells were expanded into one well of a 24 well plate (and occasionally into a 25 cm^2 flask) on gelatin to be grown up for the extraction of DNA.
10. Once confluent, the cells were harvested for DNA preparation and analysis (chapter 3).

2.1.11 Long-term Storage of ES Cells

For long-term storage, aliquots of each cell line and clone were kept at -180°C in a liquid nitrogen tank. With this method of storage, ES cells can be maintained in a viable state for many years.

2.1.11.1 Freezing ES cells.

To successfully freeze ES cells, the cells must be frozen slowly. This was performed as follows:

1. Cells were trypsinised when just confluent.
2. After spinning down, cells were resuspended in 1 ml of freezing mix consisting of foetal calf serum and 10% Dimethyl Sulphoxide (DMSO, supplied by BDH).

3. The cell suspension was placed into a 1 ml cryotube (Nunc) which was labelled with the relevant information, and placed into a polystyrene freezing box.
4. The box was left at -70 °C overnight, and the cryotube transferred to a liquid nitrogen tank. The liquid nitrogen level in the tank was checked and regularly maintained.

1.1.11.2 Thawing ES cells.

To maximise cell recovery from liquid nitrogen storage, ES cells must be thawed quickly. This was achieved as follows:

1. The appropriate cryotube was removed from liquid nitrogen storage and placed directly into a beaker of water warmed to 37 °C.
2. As soon as the cell suspension had thawed and ice crystals were no longer visible, the cryotube was swabbed with 70% ethanol and the cell suspension removed.
3. Cells were added dropwise to 10 ml of GMEM and spun down to separate the cells from the freezing mix.
4. Following resuspension in GMEM, the cells were seeded into an appropriate tissue culture vessel and grown on (section 2.1.5).

2.2 NON-ES CELL CULTURE

2.2.1 Cell Line

For the purpose of this study, Cos-7 cells were used to express a vector encoding recombinant LIF (section 2.1.3). These fibroblast-like cells are African Green Monkey kidney cells which have been transformed with an origin-defective mutant SV40 virus. Wild-type T antigen is produced by these cells making them a suitable host for transfection with vectors which require its expression. Cos-7 cells were obtained originally from American Type Culture Collection (ATCC).

2.2.2 Culture Media And Solutions

Cos-7 cells were grown in Dulbecco's modification of Eagle's medium (DMEM) supplied in powder form by Flow laboratories and made up in-house. This was supplemented with 10% heat inactivated foetal calf serum (Gibco-BRL), and the antibiotics Penicillin and Streptomycin (Gibco-BRL) at 100 units/ml and 100 µg/ml respectively. The PBS and Trypsin-EDTA used were as described in section 2.1.2.

2.2.3 Maintenance of Cos-7 Cells.

As for ES cells, all manipulations were carried out in a laminar flow cabinet. Cells were routinely grown in Nunc tissue culture flasks and incubated at 37 °C in an atmosphere of 10% carbon dioxide. Cultures were grown until just confluent and subcultured as follows:

1. Medium from a 25 cm^2 flask was aspirated and the monolayer rinsed with 5 ml PBS.
2. The PBS was aspirated, replaced with 1 ml Trypsin-EDTA and incubated for 5 minutes at 37 °C.
3. Gentle tapping was employed to dislodge the cells from the flask, which were then resuspended in 10 ml of DMEM.
4. The cells were then reseeded at 1/10 of the current density and grown on at 37 °C in an atmosphere of 10 % CO_2.

2.2.4 Freezing And Long-term Storage of Cos-7 Cell.

Cos-7 cells like ES cells, were kept for long-term storage in a liquid nitrogen tank at -180 °C. Cultures were frozen and thawed by the same methods employed for ES cells (section 2.1.11).

CHAPTER 3

MANIPULATION OF NUCLEIC ACIDS

MATERIALS AND METHODS

3.1 ISOLATION OF NUCLEIC ACIDS FROM MAMMALIAN CELLS

3.1.1 Isolation of DNA From Mammalian Cells.

3.1.1.1. DNA isolation using phenol-chloroform extraction.

This method of DNA isolation was employed in the early stages of the project. Care must be taken with this method to minimise any shearing of the DNA. All solutions were therefore mixed by gentle inversion and DNA fractions transferred using widebore pipette tips. Subsequent work utilised the quicker and less hazardous Quick lysis method.

1. Cells were harvested as described in 2.1.5 and the cell pellet of a 25 cm^2 flask resuspended in 1 ml of lysis buffer containing 150 mM Sodium Chloride, 100 mM EDTA, 100 mM Tris-HCl (pH 8.0) and 0.5% Sodium dodecyl sulphate (SDS).
2. RNA was removed by adding 10 µg RNase A (Sigma) per ml of cell suspension and incubating at 37°C for one hour.
3. Proteinase K (Boehringer Mannheim) was added to give 100 µg/ml and the cell lysate incubated at 50°C for three hours.
4. After incubation, an equal volume of water-saturated Phenol (Boehringer Mannheim) was added and the solutions mixed by gentle inversion until an emulsion was obtained.
5. The two phases of the emulsion were separated by centrifugation at 3000 rpm (Sorvall Coolspin) for ten minutes.
6. The aqueous (top) phase was removed using a wide bore pipette tip and placed into a clean tube.
7. This aqueous phase was extracted twice more, once with 0.5 volume phenol + 0.5 volume chloroform, and then with an equal volume of a 24:1 mixture of chloroform-isoamylalcohol.

8. After the last extraction, the aqueous phase was placed into a clean tube and precipitated with 0.5 volume 7.5 M Ammonium acetate and three volumes of 100% Ethanol.
9. Precipitated DNA was hooked out using a sterile glass rod and rinsed in 70% Ethanol.
10. The DNA was briefly air dried (approximately two minutes) and then placed into 500 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.4) and left to dissolve overnight at 4°C.

3.1.1.2 DNA isolation from mammalian cells using quick lysis technique.

This technique is described by Laird *et al.* (1991) and was chosen as the preferred method of DNA isolation because it was simple, quick, high yielding, and involved minimal shearing forces on the DNA. In addition, this method did not require the use of Phenol, and produced DNA which was of sufficient quality to be consistently digested by restriction enzymes.

1. Cells were harvested as described in 2.1.5 and spun down. The pellet was resuspended in 5 ml of quick lysis buffer (for a 25 cm^2 flask), which contains 100 mM Tris-HCl pH 8.5, 5 mM EDTA, 0.2% (w/v) SDS, 200 mM Sodium chloride and 100 µg Proteinase K (Boehringer Mannheim).
2. The cell lysate was incubated at 37°C overnight and the DNA precipitated with an equal volume of isopropanol.
3. Using a glass rod, the DNA precipitate was hooked out and washed briefly in 70% ethanol.
4. After air drying for two minutes, the DNA was placed into 500 µl TE buffer and left at 4°C overnight to dissolve.
5. To ensure that the DNA was completely dissolved, the solution was incubated at 50°C for two hours.

3.1.2 Quantification Of DNA.

The concentration of DNA was estimated by reading its optical density (OD) at 260 nm. A 1/100 dilution was usually made in distilled water and read against a blank of distilled water. After correction for dilution, an OD of 1.0 is equivalent to 50μg/ml of double stranded DNA.

3.2 BACTERIAL CELL CULTURE

All targeting vectors and the mammalian expression vector pC10-6R (for LIF production) were constructed in bacterial plasmids which, upon transformation into a suitable host strain, could be replicated along with the host DNA thereby amplifying the plasmid.

3.2.1 Bacterial Host Strains.

JM83 F-,araΔ(lac-proAB),rspL,(Strr),(φ80dΔ(lacz)M15) (Viera and Messing 1982). This strain was used for propagation the plasmid pC10-6R.

XL1-Blue Δ(mcrA)182, Δ(mcrCB-hsdSMR-mrr)172, endA1, supE44, thi-1, recA, gyrA96, relA1, lac, λ-, [F', proAB, $lac^{q}ZΔ$ M15, Tn10, (tet^{r})] (Bullock *et al.* 1987). This strain was used for propagation of the 'Hit and Run' targeting vectors.

3.2.2 Media and Additives

Terrific broth (TB)

12 g Bacto-tryptone (Difco)

24 g Yeast extract (Difco)

4 g Glycerol (Sigma)

Distilled water to 900 ml.

After autoclaving, 100 ml of sterile phosphate solution (0.1 M Potassium dihydrogen phosphate and 0.72 M diPotassium hydrogen phosphate) was added.

L-Broth Agar

2.46 g Magnesium sulphate

10 g Tryptone

5 g Yeast extract

10 g Sodium chloride

15 g agar (Oxoid Ltd)

Distilled water to 1 litre.

All medium was sterilised by autoclaving at 121°C for 20 minutes.

Ampicillin

The antibiotic ampicillin was added to medium to select for transformed bacteria which express an ampicillin resistance gene carried by the plasmid. A stock solution of ampicillin (Sigma) was made up in distilled water at a concentration of 50 mg/ml. This was filter sterilised and stored at -20°C until required, when it was added to medium to give a final concentration of 50 μg/ml.

3.2.3 Preparation Of Competent Cells For Heat Shock Transformation.

This is the method developed by Hanahan (1983) and yields bacteria which are competent for transformation at high efficiency by heat shock.

1. A fresh, single colony of JM83 was used to inoculate a 5 ml overnight culture of L-Broth which was shaken at 37°C.
2. The next day, 2 ml of the overnight culture was used to inoculate 100 ml of prewarmed L-broth, and incubated with shaking at 37°C until an $OD_{550\ nm}$ of 0.6 was achieved.
3. At this point, the culture was transferred to a pre-chilled, sterile flask and incubated on ice for 10 minutes.
4. The culture was split between two sterile 50 ml Falcon tubes, and spun down at 3000 rpm (Sorvall RC-5B, DuPont Instruments) at 4°C for 5 minutes.

5. The media was carefully removed and the pellet resuspended in 20 ml of ice-cold transformation buffer (30 mM Potassium acetate, 100 mM Rubidium chloride, 10 mM Calcium chloride, 50 mM Manganese chloride, 15% Glycerol (v/v), pH 5.8).
6. The cell suspension was cooled on ice again for 10 minutes and then spun down and pelleted again.
7. After removal of the supernatant, the cell pellet was suspended to a final volume of 2 ml in a buffer containing 10 mM PIPES, 75 mM Calcium chloride, 10 mM Rubidium chloride, 15% Glycerol (v/v), pH 7.0.
8. The cell suspension was cooled on ice for 15 minutes and then aliquoted (50 µl) into precooled tubes and snap-frozen in a dry ice-ethanol bath.

3.2.4 Heat Shock Transformation Of Bacterial Cells.

1. The plasmid to be transformed was diluted to 50 ng in 5 µl in sterile, distilled water.
2. Competent JM83 cells were thawed on ice and 50 ng (5 µl) of the plasmid to be transformed was added and mixed with the cells.
3. After chilling on ice for 30 minutes, the cells were heat shocked by placing them in a waterbath at 42°C for exactly 90 seconds and then returning them back onto ice for 2 minutes.
4. L-Broth was added to make the cell suspension up to 1 ml, which was then incubated on a shaker at 37°C for one hour.
5. After one hour, the antibiotic resistance genes carried by the plasmid should be expressed in transformed cells These were selected by plating 100 µl of the cell culture onto L-agar plates containing ampicillin.
6. Transformed cells were visible as colonies after overnight incubation at 37°C.

3.2.5 Small Scale Plasmid Preparation.

This is a quick method of isolating small amounts of crudely purified plasmid DNA from transformed cells, and was an adaptation of the method described by Birnboim and Doly (1979).

1. An overnight culture of cells was set up in 10 ml of Terrific broth.
2. A small amount of the culture (approximately 1.5 ml) was removed and spun down at 12 000 g in a microcentrifuge.
3. The supernatant was removed and the pellet resuspended in 100 µl of a solution of 50 mM Glucose, 25 mM Tris-HCl (pH 8.0), 50 mM EDTA, 10 mg/ml Lysozyme (Sigma).
4. After standing at room temperature for 10 minutes, 200 µl of a solution of 1% (w/v) SDS and 0.2 M Sodium hydroxide was added to the cell lysate. The solutions were gently mixed and left on ice for 5 minutes.
5. Then, 150 µl of Potassium acetate was added and the mix left for a further 5 minutes on ice.
6. The cell lysate was spun down at 12 000 g in a microcentrifuge and 400 µl of the supernatant removed to a clean tube.
7. To precipitate the DNA, 240 µl (0.6 volumes) of isopropanol was added, and the supernatant left at room temperature for 5 minutes.
8. The DNA was spun down at 12 000 g and the supernatant removed.
9. The pellet was washed with 70% ethanol, dried under vacuum for 5 minutes, and then resuspended in 30 µl of TE buffer.

3.2.6 Large Scale Plasmid DNA Preparation.

This method of plasmid DNA preparation was used to produce large quantities of high quality plasmid DNA suitable for use in the transfection of mammalian cells, and is as described in Sambrook *et al.* (1989).

1. A 500 ml overnight culture in Terrific Broth was spun down at 6000 rpm (Sorvall RC-5B, DuPont Instruments) for 5 minutes.
2. The supernatant was removed and the pellet resuspended in 20 ml of a solution containing 50 mM Glucose, 25 mM Tris-HCl, 50 mM EDTA, and 10 mg/ml Lysozyme (Sigma).

3. After standing at room temperature for 5 minutes, 40 ml of a solution of 1% (w/v) SDS and 0.2M Sodium hydroxide was added, and the lysate left on ice for 10 minutes.
4. A volume of 30 ml of pre-cooled Potassium acetate was added, and the lysate mixed gently by inversion. This was incubated on ice for a further 20 minutes.
5. Cell debris was pelleted by centrifugation at 12 000 rpm (Sorvall RC-5B, DuPont Instruments), 4° C for 30 minutes.
6. The resulting supernatant was filtered through muslin into a clean vessel and mixed with 0.6 volumes of isopropanol.
7. After standing at room temperature for 5 minutes, the DNA was spun down at 10 000 rpm (Sorvall RC-5B, DuPont Instruments) at room temperature, for 30 minutes in a prewarmed rotor.
8. After discarding the supernatant, the pellet washed in 70% ethanol and air dried.
9. When dry, the pellet was dissolved in 7.7 ml of TE buffer to which 400 µl of 10 mg/ml Ethidium bromide and 9.2 g of Caesium chloride was then added.
10. The resultant solution was divided between two quickseal ultracentrifuge tubes (Beckmann). These were balanced to within 0.01 g of each other, heat-sealed, and centrifuged in a fixed-angle rotor ultracentrifuge (Beckmann TL-100) at 100 000 rpm for 16-18 hours at 20 °C.
11. Circular plasmid DNA was visible as a tight band approximately half way up the centrifuge tube, and was extracted using a sterile 0.8 gauge hypodermic needle (Becton Dickinson) attached to a syringe.
12. Removal of the Ethidium bromide from the DNA solution was achieved by repeated extractions with water-saturated butan-1-ol until the organic phase became colourless.
13. The DNA was then precipitated from this solution with 2.5 volumes of 70% ethanol and either hooked out on a glass rod, or spun down by centrifuging at 10 000 rpm (Sorvall RC-5B, DuPont Instruments) for 10 minutes.
14. After washing briefly in 70% ethanol, the DNA was dried under vacuum and resuspended in 200 µl of TE buffer.

3.3 ANALYSIS OF DNA BY RESTRICTION ENZYME DIGESTION

Digestion of DNA by restriction enzymes was carried out under the conditions recommended by the manufacturer (usually Boehringer Mannheim) using the appropriate buffer (supplied by manufacturer). When two enzymes were to be used, the digests were carried out simultaneously if the buffer was compatible with both. Otherwise, digestion with the enzyme requiring the lower salt concentration was carried out first and then the salt concentration was altered to that suitable for the second enzyme by addition of Sodium chloride. Restriction digests were terminated by heat inactivation at 80° C for 10 minutes, or if to be run out on an agarose gel, by the addition of 1/10 volume of stop mix (100 mM EDTA pH 8.0, 20% (w/v) Ficoll (Pharmacia), and Orange G (Sigma).

3.3.1 Plasmid DNA Digest.

Typically, 1 µg of plasmid DNA was digested with one unit of restriction enzyme in a total volume of 20 µl. Incubation was at the recommended temperature (usually 37° C), for 1-2 hours.

3.3.2 Genomic DNA Digest.

For digests of genomic DNA, 6-8 µg of DNA was digested with 20 units of restriction enzyme in a total volume of 40 µl. Bovine serum albumin (BSA), supplied by Boehringer Mannheim, was added to give a concentration of 0.1 mg/ml if the DNA had been prepared by the 'Quick lysis' method. If a high salt buffer was used, Spermidine (Sigma) was added to a concentration of 5 mM. Incubation was overnight at a temperature recommended for the enzyme.

3.4 GEL ELECTROPHORESIS OF NUCLEIC ACIDS

Fragments of DNA and RNA were size separated through horizontal agarose gels by electrophoresis.

3.4.1 Solutions And Buffers.

10 X TBE

890 mM Tris-HCl

20 mM EDTA pH 8.0

890 mM Boric acid pH 8.3

20 X TAE

0.8 M Tris-HCl

20 mM EDTA pH 8.0

0.4 M Acetic acid

10 X DNA loading buffer ('Stop mix')

20 % (w/v) Ficoll (Pharmacia)

100 mM EDTA

Orange G (Sigma) to colour.

3.4.2 Agarose Gel Electrophoresis Of DNA.

Separation of DNA molecules according to size was achieved by electrophoresis through an agarose gel of the appropriate density. Density was varied by altering the percentage of agarose (Type II medium EEO, Sigma) in the gel.

Restricted genomic DNA was run slowly at 20 Volts overnight in a 0.8-1.0 % gel made up in 1 X TAE buffer. Continuous circulation of buffer was maintained by use of a peristaltic pump. Smaller DNA fragments, such as plasmid DNA or PCR products were run in 1-2 % agarose gels, and fragments smaller than 400 bp were resolved either by running in a 4 % Nusieve (FMC Bioproducts) gel, or a gel containing 3 % Nusieve and 1 % agarose These were run out more quickly than genomic DNA, usually between 60-100 Volts for up to one hour. For these gels, 1 X TBE buffer was used. Ethidium bromide was added to all buffers at a concentration

of 1 µg/ml. Stop mix of 1/10 sample volume was added to all samples prior to loading onto the gel.

Molecular weight markers were run alongside DNA samples to facilitate band sizing. When separating fragments of 1-10 kb, Hind III digested lambda DNA (BRL) was used, for small fragments under 1 kb e.g. PCR products, Hae III digested ϕX DNA (BRL) was used.

Electrophoresed DNA was visualised on a UV transilluminator (UVP Products Ltd) at 305 nm, and photographed using a video copy processor (Mitsubishi).

3.4.3 Preparative DNA Electrophoresis.

Specific DNA fragments to be used as probes for radiolabelling were isolated using gel electrophoresis followed by processing with one of the two following methods:

3.4.3.1 DNA probe preparation using 'Geneclean'.

A restricted DNA sample was electrophoresed through an agarose gel of the appropriate density in 1 X TBE buffer. The band to be used as a probe was carefully excised from the gel and extracted from the gel slice using the 'Geneclean' reagents (BIO 101) and protocol. The DNA was diluted in sterile water to give a concentration of 30 - 50 ng/µl.

3.4.3.2 Preparation of DNA probe using low melting point agarose.

The restricted DNA to be used as a probe was run out in a low melting point agarose gel (Ultrapure LMP agarose, BRL) and excised as a gel slice. The DNA was not isolated from this slice, but instead the slice was weighed and diluted with sterile water in a ratio of 3 ml/g of gel, so that 11 µl of diluted gel slice contained 25-50 ng of DNA.

3.5 TRANSFER OF NUCLEIC ACID TO MEMBRANES

3.5.1 Southern Transfer Of DNA.

The transfer of DNA form agarose gel to nylon membranes was achieved using an adaptation of the capillary blotting method of Southern (1975).

1. Following electrophoresis, gels were placed upon a UV transilluminator, aligned with a ruler (to facilitate future fragment sizing), and photographed.
2. DNA within the gel was denatured by gentle agitation of the gel in a solution of 0.5 M Sodium hydroxide and 1.5 M Sodium chloride, for 45 minutes.
3. The gel was then neutralised by further agitation in a solution of 1 M Tris-HCl and 2 M Sodium chloride, pH 5.5, for 45 minutes.
4. A large strip of 3 MM (Whatman) filter paper was soaked in 20 X SSC (3 M Sodium chloride, 0.3 M TriSodium citrate, pH 7.4) and placed on a narrow board.
5. A wick was generated by placing the board over a reservoir of 20 X SSC and placing the ends of the filter paper into the reservoir.
6. The gel was laid onto this wick and exposed regions of the wick were then covered with Saran wrap (Dow Chemical Company).
7. A piece of nylon membrane (Hybond-N, Amersham) was cut to the exact size of the gel and placed onto it. In order to eliminate any air bubbles which would interfere with the transfer of DNA, a 10 ml glass pipette was rolled gently over the hybond.
8. The hybond was then covered with a piece of 3 MM filter paper (also cut to size), which had been prewetted in 2 X SSC.
9. A further two pieces of 3 MM filter paper (not prewetted this time) were added followed by a stack of absorbent paper towels to a height of 12 cm and a glass plate. A weight of approximately 1 kg was applied to the top.
10. The blot was left to transfer for a suitable amount of time, which, for genomic blots, was no less than four hours and usually overnight.
11. When transfer had occurred, the blot was dismantled and the membrane washed briefly in 2 X SSC to remove agarose fragments.

12. The DNA was bound to the membrane, first by UV-mediated crosslinking by exposure to 1200 μJ UV irradiation in a Stratalinker (Stratagene), followed by baking at 80°C for 30 minutes.

3.6 RADIOLABELLING OF DNA

3.6.1 Random Primed Labelling of DNA.

This method of random primed DNA labelling is based upon that of Feinberg and Vogelstein (1983), in which a mixture of all possible hexanucleotides are hybridised to the DNA to be labelled. A complementary strand is synthesised using the 3'OH termini of the random hexanucleotide primer and labelling grade Klenow enzyme. Radiolabelled [α-^{32}P] dCTP is incorporated into the newly synthesised DNA as the only source of dCTP, producing an equal degree of labelling along the entire length of input DNA which is not influenced by the length of DNA to be labelled.

1. The DNA to be labelled was diluted with sterile distilled water to give 25-50 ng of DNA in a volume of 11 μl.
2. This was denatured by heating at 100°C for 10 minutes, and then placed immediately onto ice unless the probe originated from a gel slice, in which case it was cooled to 37°C.
3. Reagents from a Boehringer Mannheim random prime labelling kit were added as follows:

   2 μl 10 X reaction buffer

   1 μl each of dATP, dGTP, dTTP, (0.5 mM)

   2 Units of Klenow enzyme

   30 μCi [α-^{32}P] dCTP (Amersham)
4. The reaction mix was incubated at 37°C for 45 minutes after which time the level of radiolabel incorporation was measured by Trichloroacetic acid precipitation (Sambrook *et al.* 1989). A sufficiently high level of incorporation was deemed to be 40 % of the radiolabel.

5. Unincorporated nucleotides were removed by running the probe through a Nick column (Pharmacia) and eluting in a volume of 400 µl of TE buffer.
6. To minimise non-specific binding of the probe, 500 µg sonicated Salmon sperm DNA (Sigma) was added to the probe DNA, which was then denatured by heating at 100 °C for 10 minutes.
7. The probe was then added to the hybridisation mix.

3.6.2 Pre-annealing of Repetitive Sequences.

The repetitive elements contained within probes were annealed after labelling to prevent the repeats taking part in the hybridisation reaction. The procedure employed was adapted from Sealey *et al.* (1985).

1. Labelling was carried out as described in section 3.6.1 but the probe was not run through a Nick column.
2. Instead, 1 mg of sonicated human DNA was added to the probe along with 500 µg of sonicated Salmon sperm DNA, and denatured by heating at 100 °C for 10 minutes.
3. To allow annealing of the repetitive sequences, the DNA was incubated at 68 °C for 30-45 minutes.
4. After incubation, the probe was added to the hybridisation mix.

3.6.3 End Labelling Of DNA Oligonucleotides

Oligonucleotides were labelled by the transfer of the ^{32}P-labelled γ-phosphate from [γ-^{32}P] ATP by the action of the enzyme T4 polynucleotide kinase (T4 PNK), onto the terminal 5'-OH group. A reaction mix in a 20 µl volume was set up as follows:

30 ng Oligonucleotide
1 X Polynucleotide Kinase (PNK) buffer (5 mM Tris-HCl pH 8.0, 1 mM Magnesium chloride, 0.5 mM Dithiothreitol).
10 Units Polynucleotide Kinase (Boehringer Mannheim)
50 µCi [γ-^{32}P]ATP (Amersham)

The reaction was incubated at 37 °C for 30 minutes and then added to the hybridisation mix.

3.7 NUCLEIC ACID HYBRIDISATION

3.7.1 Hybridisation Solutions.

Hybond hybridisation mix.

6 X SSC (20 X SSC is 3 M Sodium chloride and 0.3 M TriSodium citrate, pH 7.4)

10 % (w/v) Dextran sulphate

0.1 % (w/v) Sodium pyrophosphate (Sigma)

0.5 % (w/v) SDS

4 X Denhardts solution (100 X Denhardts contains 2% (w/v) BSA (Sigma), 2 % (w/v) Polyvinylpyrrolidone, 2 % Ficoll (Pharmacia), 1 mM EDTA)

Oligonucleotide hybridisation mix.

5 X SSC

0.05 % (w/v) BSA

0.05 % (w/v) Ficoll

0.1 % (w/v) SDS

0.05% (w/v) Polyvinylpyrrolidone

0.1 % (w/v) Sodium Pyrophosphate

3.7.2 Prehybridisation And Hybridisation Protocols.

Prehybridisation of filters in hybridisation mix containing denatured, sonicated Salmon sperm DNA was carried out before addition of the probe to block any non-specific binding sites.

1. Filters were placed between two gauze sheets (which are used to prevent drying out of the filter during hybridisation) and air bubbles were eliminated by submerging and rolling these up under a solution of 2 X SSC.

2. The filter-gauze roll was placed into a Hybaid hybridisation bottle and unrolled in a solution of 2 X SSC.
3. This was poured off and replaced with the appropriate prewarmed hybridisation mix, (Hybond hybridisation mix for random primed-labelled probes, or Oligonucleotide hybridisation mix for end-labelled probes). A volume of 10 ml was used for small hybridisation bottles, and 20 ml for large bottles.
4. To this, 500 µg of denatured, sonicated Salmon sperm DNA was added, the bottle sealed and placed into the rotating spindle of a Hybaid hybridisation oven.
5. Prehybridisation was carried out for at least one hour at 5 °C below the T_M of the oligonucleotide (see section 3.9.1) for end labelled oligonucleotides, or at 68 °C for random prime-labelled probes.
6. Following prehybridisation, the probe was added and hybridised to the filter overnight at the same temperature used for the prehybridisation.

3.7.3 Post-hybridisation Washing Protocols.

Following overnight hybridisation, the filters were washed to remove background (non specific) hybridisation of the probe. The initial wash solution contained 2 X SSC, 0.1 % (w/v) SDS and 0.1 % (w/v) Sodium pyrophosphate, heated to the hybridisation temperature. Filters were washed at 68°C with gentle agitation for twenty minutes after which time the background radioactivity was monitored. If the background count was still high, washes of increasing stringency (decreasing SSC concentration) were employed. The most stringent wash employed was 0.1 X SSC.

3.7.4 Removal Of Hybridisation Signal.

Removal of the hybridisation was necessary for the reprobing or storage of filters, and was achieved by one of the following two methods. For successful removal of probes, it is important never to let the filter dry out.

3.7.4.1 Removal of hybridisation signal by the alkaline stripping method.

Removal of hybridisation signal was achieved by incubating the filter at 45°C for 30 minutes in a solution of 0.4 M Sodium hydroxide with gentle agitation. This solution

was then replaced with a solution of 0.1 X SSC, 0.1% (w/v) SDS and 0.2 M Tris-HCl, pH 7.5, and incubated for a further 30 minutes.

3.7.4.2 Removal of hybridisation signal by the boiling method.

With this method, removal of signal was achieved by adding the filter to a boiling solution of 0.1% (w/v) SDS and allowing it to cool to room temperature.

Complete removal of hybridisation signal was verified by use of one of the methods for hybridisation signal detection described in section 3.8.

3.8 DETECTION OF HYBRIDISATION

3.8.1 Autoradiography.

Hybridised filters were wrapped in Saran wrap to prevent drying out and exposed to Kodak X-OMAT film in light-proof cassettes with intensifying screens. Filters were exposed to film at a temperature of -70 °C for periods of 30 minutes to 3 weeks, depending upon signal strength. The film was developed using a Fuji RG II automatic X-ray film processor.

3.8.2 Phosphorimaging.

A more rapid visualisation of hybridisation signal was achieved by use of a phosphorimage system. The hybridised filters were wrapped in Saran wrap, placed into a phosphorimage cassette (Molecular Dynamics) and exposed to the phosphor screen at room temperature for a period of between 30 minutes and 36 hours, depending upon signal strength. After an appropriate exposure time, the phosphorimage screens were scanned on a PhosphorImager (Molecular Dynamics), where a laser beam converts the radioactive signal into a digital image. The amount of radioactivity was represented by a proportional variation in pixel values, and the image was displayed on a grey scale. Good visualisation of the image was achieved by adjustment of the upper and lower limits of the grey scale, and the image was printed out on a grey scale laser printer.

3.9 AMPLIFICATION OF DNA BY THE POLYMERASE CHAIN REACTION (PCR)

The Polymerase chain reaction (PCR) is an enzymatic amplification technique which permits the cyclic amplification of a specific DNA sequence (Saiki *et al.* 1985, Mullis and Faloona, 1987). Sequence specificity is provided by a pair of oligonucleotide primers which direct amplification and are complimentary in sequence to the 5' ends of the sequence to be amplified. The primers anneal to the template DNA under conditions permitting only exact complimentary annealing, and prime the extension of complimentary DNA by the enzyme DNA polymerase.

The PCR reaction is a series of cycles consisting of three steps each involving an incubation at different temperatures for set periods of time. The first step is a high temperature incubation causing denaturation of the template DNA, which is necessary for its amplification. This is followed by rapid cooling to a temperature which promotes precise annealing of the oligonucleotide primers. A period at 72°C follows, this is the optimum temperature for extension of DNA by the thermostable DNA polymerase.

3.9.1 PCR Of Genomic DNA.

3.9.1.1 PCR setup.

A PCR reaction was set up in a 50 µl volume as follows:

100 ng template DNA

5 µl 10 X PCR buffer II (Perkin Elmer)

3 µl (25 mM) Magnesium chloride (Perkin Elmer)

1 µl each primer (250 ng/µl)

10 mM each dNTP (Pharmacia)

1 unit thermostable AmpliTaq DNA polymerase (Perkin Elmer)

The reaction mix was overlaid with mineral oil (Sigma) to prevent evaporation, and placed onto a Hybaid Omnigene thermal cycler. The temperature and length of each

step was varied according to the primer pair and the length of the fragment to be amplified.

Each PCR reaction always included an initial cycle with a long denaturation step at a higher temperature, usually 94 °C for 3 minutes. This was followed by a set of cycles, the number of which was dependent on the efficiency of the amplification (usually 30). This set consisted of a denaturation step (usually 92 °C for 30 seconds), followed by an annealing step. The annealing temperature was assessed for each primer pair and was taken as 5 °C lower than the melting temperature (T_M) of the primers. The T_M was determined by the equation T_M (at 1 M Na^+ = 4(G+C) + 2(A+T). Annealing was generally carried out for one minute and was followed by extension at 72 °C for approximately one minute/kb to be amplified. The final cycle incorporated an increased extension step, usually of 7 minutes.

The products of a PCR reaction were visualised by running 10 µl on an ethidium bromide -stained agarose gel (section 3.4).

3.9.1.2 Primer sequences.

PRIMER	SEQUENCE	SIZE (bp)
M10AI	CAGAGGGAATTATTAAGCACAGTGG	25
M10ΔF	GGTACTATCAAAGAAAATATCATCGG	26
MI10	GGTTGACACAAGTTTCTAGGATAGC	24
M10AI	CAGAGGGAATTATTAAGCACAGTGG	25
REVERSE	AACAGCTATGACCATG	16

3.9.1.3 PCR conditions.

5' PRIMER	3' PRIMER			PROGRAM		PRODUCT
		SEGMENT	TEMP (°C)	TIME (secs)	No. CYCLES	SIZE (bp)
M10ΔF	**MI10**	STEP 1	94	60	1	800
		STEP 2	45	30	35	
			72	60		
			92	30		
		STEP 3	45	30	1	
			72	180		
M10AI	**MI10**	STEP 1	94	60	1	1000
		STEP 2	55	30	35	
			72	60		
			92	30		
		STEP 3	55	30	1	
			72	180		
M10ΔF	**REVERSE**	STEP 1	94	60	1	900
		STEP 2	45	30	35	
			72	60		
			92	30		
		STEP 3	45	30	1	
			72	180		

CHAPTER 4

TRANSGENIC TECHNIQUES

MATERIALS AND METHODS

4.1 Technical Considerations.

Embryonal stem cells, when returned to the environment from which they were originally isolated, can behave like normal cells of the blastocyst inner cell mass, and contribute to all of the tissues of the resulting animals. To retain this ability, ES cells must be maintained in a pluripotent state (section 2.1).

In this study, we attempted to make chimaeras from targeted cell lines by the coculture/aggregation method (Nagy and Rossant 1993). This technique involves the coculture of ES cells and morulae resulting in the internalisation of the ES cells into the developing embryo. The morulae are cultured overnight to the blastocyst stage and implanted into a pseudopregnant foster mother, where they go on to form viable progeny which should include a contribution from the introduced ES cells. This method is thought to be a quicker and simpler than the alternative, technically demanding, method of microinjection (Gordon *et al.* 1980). The microinjection technique usually requires many hours of practice before the necessary level of manual dexterity is achieved, and in contrast to the aggregation technique, utilises elaborate and expensive equipment (Wood *et al.* 1993). The aggregation technique does however, have disadvantages associated with it. The zona pellucida is removed prior to aggregation, requiring that the morula and ES cells be cultured overnight until a blastocyst is formed. Both ES cells and the morulae have differing nutritive requirements at this stage, consequently the culture media employed for the overnight incubation represents a compromise between the two requirements and is optimal for neither cell type.

The day to day husbandry of the mice involved in this study as well as administration of the hormones required for the superovulation regime, was carried out by staff of the Biomedical Research Facility.

4.2 Preparation of ES Cells For Chimaera Production.

ES cells for chimaera production were maintained under optimal conditions as far as possible. Cells were passaged at least once after thawing as described in section 2.1.

On the day of aggregation, a subconfluent culture of exponentially dividing cells was washed with 5 ml of PBS and incubated with 1 ml of half-strength trypsin-EDTA. As soon as the cells formed loose clumps, 10 ml of GMEM was added to inactivate the trypsin. The cell suspension was left in the flask at 37 °C and used for aggregation within one hour.

4.3 Mouse Strains.

A F_1 hybrid mouse strain from a C57BL/6 X CBA cross (C57BL/6 being the female) was used to generate morulae for chimaera production. These mice display hybrid vigour, are a compatible host for 129/Ola ES cells, and carry pigmentation and other marker alleles which differ sufficiently from those carried by the ES cells to facilitate early detection of chimaeras. The strain used for foster mothers and vasectomised stud males was the Swiss albino mouse strain. An albino strain was used for this purpose so that in the eventuality of any embryos being generated by the foster mothers (e.g. if the vasectomised stud male produces sperm, or a non-vasectomised stud male is used in error), the offspring would be easily detected as non-pigmented albinos within a pigmented litter.

4.4 Superovulation of Females For Embryo Collection.

In order to collect a large number of morulae from a single mouse, female mice were induced to superovulate by administration of Pregnant Mares Serum Gonadotrophin (PMSG, Intervet). This stimulates follicles in the ovary to mature and was given as an intra-peritoneal (i.p.) injection of 5 i.u. (international units). This was followed 48 hours later by i.p. injection of 5 i.u. of Human Chorionic Gonadotrophin (HCG), which overrides the mouse's own luteinising hormone and stimulates the mature follicles to ovulate. The females were set up for mating with stud males 2-6 hours after HCG injection and were checked for vaginal plugs the following morning. The presence of a vaginal plug was taken as evidence of mating.

4.5 Preparation of Pseudopregnant Foster Mothers.

Pseudopregnant foster mothers at an appropriate stage of pregnancy were required to provide a uterine environment which was receptive to embryo implantation and development. Foster mothers whose pseudopregnancies were at a stage compatible with that of the transferred blastocysts, were prepared by mating female mice in oestrus (identified by appearance of vaginal epithelium) to vasectomised, sterile males. The stimulus of mating causes the female to display the hormonal profile of a pregnant female, and subsequent embryo transfer at the appropriate time provides a signal for further maintenance of pregnancy. Blastocysts, at 3.5 d. p. c., were transferred into pseudopregnant females at 2.5 d. p. c. of pseudopregnancy. This discrepancy is to allow the blastocysts to resynchronise with the uterus and recover from a period of culture *in vitro*.

4.6 Medium and Solutions.

M16 Medium

This is a modified Krebs-Ringer bicarbonate salt solution as described by Whittingham (1971). It was prepared in-house in 100 ml quantities from the individual components listed below, filter sterilised through a 0.2 μm filter (Gelman), and stored at 4 °C for up to one month. Morulae and ES cells were cultured in microdrops (25-50 μl) of this medium under a layer of filtered, light paraffin oil (BDH), at a temperature of 37 °C in an atmosphere of 5% CO_2. The microdrops were prepared in tissue culture plates and equilibrated in a CO_2 incubator for at least 30 minutes prior to use.

0.553 g Sodium chloride
0.036 g Potassium chloride
0.025 g Calcium chloride dihydrate
0.016 g Potassium dihydrogen phosphate
0.029 g Magnesium sulphate hexahydrate
0.210 g Sodium bicarbonate (Sigma)
0.32 ml (60% (w/w) syrup) Sodium lactate (Sigma)
0.004 g Sodium pyruvate (Sigma)

0.1 g Glucose

0.4 g BSA (Sigma)

0.006 g Penicillin G Potassium salt (Sigma)

0.005 g Streptomycin sulphate (Sigma)

0.001 g Phenol red

All made up to 100 ml with ultrapure distilled water.

M2 Medium

This medium is similar to M16 medium but has some of the bicarbonate replaced by HEPES buffer. It too was prepared in house in 100 ml quantities from individual components listed below, and filter sterilised through a 0.2 μm filter (Gelman). This medium was stored at 4 °C for up to one month. All work involving the handling of morulae outside of the incubator was carried out in M2 medium.

0.553 g Sodium chloride

0.036 g Potassium chloride

0.025 g Calcium chloride

0.016 g Potassium dihydrogen phosphate

0.029 g Magnesium sulphate

0.035 g Sodium bicarbonate

0.497 g HEPES

0.32 ml Sodium lactate (60% syrup)

0.004 g Sodium pyruvate

0.1 g Glucose

0.4 g BSA

0.006 g Penicillin G

0.005 g Streptomycin sulphate

0.001 g Phenol red

Made up to 100 ml with ultrapure distilled water.

4.7 Morula Collection

Eight cell stage embryos for aggregation were recovered from superovulated females as follows:

1. Superovulated females were sacrificed at 2.5 d. p. c.
2. The oviducts and upper part of the uterus were dissected out, any body fat trimmed away, and placed into a drop of M2 medium.
3. Morulae were washed out from the oviducts by flushing through approximately 0.5 ml M2 medium with the aid of a 30 gauge needle, or released by tearing open the oviducts using fine forceps.
4. A mouth controlled pipette attached to a drawn out glass Pasteur pipette was used to collect all the liberated morulae and transfer them to microdrops of M2 medium.

4.8 Aggregation Formation.

Embryo-ES cells aggregates were prepared with the aid of Phytohaemagglutinin (Sigma) which helps stick the ES cells to the embryo. Clumps of ES cells were stuck to single morulae or sandwiched between a pair of embryos. There seems to be some debate as to whether a higher rate of chimaera production is obtained from aggregates constructed with a single embryo, or with a pair of embryos, and for this reason we constructed both. Bradley (1987) suggests that a single embryo might not be able to properly regulate a group of ES cells, whereas a combination of two embryos might overcome this problem. However, Nagy (1994) proposes that a higher percentage of germ line transmission might be achieved from chimaeras resulting from aggregations of single embryos with ES cells. He argues that in 50 % of single embryo aggregations the favourable combination of XX (female) morula and XY (male) ES cells can occur, compared to only 25 % of two embryo sandwich aggregations. This combination is desirable as the resulting chimaera will be phenotypically male with the XX primordial germ cells derived from the host embryo unable to form gametes. Therefore germline contribution of the ES cells is favoured.

Embryo and ES cell aggregates were constructed as follows:

1. Removal of the zona pellucida (a necessary prerequisite for aggregation with ES cells) was achieved by transferring the embryos into a 0.5 % solution of Pronase (Sigma) in M2 medium for approximately five minutes.
2. Once the zona pellucida had disappeared, the embryos were passed through two microdrops of M2 medium to wash off excessive pronase, and transferred in batches of approximately 10, to a microdrop of 0.2 µg/ml Phytohaemagglutinin (Sigma) in M2 medium.
3. A small amount of ES cell aggregates were transferred into the same microdrop, and aggregates prepared by sticking together clumps of ES cells containing 5-10 cells to single morula, or by sandwiching them between two morulae.
4. Aggregates were washed through two drops of M2 medium to remove traces of the Phytohaemagglutinin, and placed into individual drops of M16 medium.
5. After overnight incubation, the majority of aggregates had formed blastocysts and were ready for implantation.

4.9 Transfer of Manipulated Embryos To Pseudopregnant Host Females.

Manipulated embryos were transferred into pseudopregnant females whose stage of pregnancy was one day behind that of the embryos. At this stage of pregnancy the blastocysts would normally be found in the uterus and therefore this is where the manipulated embryos were returned. The uterine environment of the pseudopregnant females should be receptive to implantation of these embryos and expedite their development into viable offspring. This surgical procedure was usually carried out on the afternoon of the day following morula aggregation, and was performed on the open bench under general anaesthesia as follows:

1. Blastocysts to be transferred were collected together into a microdrop of M2 medium.
2. The transfer pipette (a drawn out Pasteur pipette attached to a mouth pipette) was prepared by taking up liquid paraffin to the length of the drawn out section, followed by a small amount of air (to create an air bubble), a small amount of M2

medium, and another air bubble. The embryos to be transferred were taken up next, in as small an amount of medium as possible, followed by another air bubble and a small amount of medium. Paraffin oil provided a means of slowing down the movement of liquids when expelled from the pipette, as too forceful a transfer could result in the embryos being washed straight through the uterus and out of the cervix. The air bubbles were incorporated as a visual marker to indicate when the medium containing the embryos had passed into the uterus.

3. Swiss albino foster mothers were anaesthetised by administration of an i.p. injection of a solution containing 0.75 ml Hypnorm analgesic (Roche), 0.5 ml Hypnovel anaesthetic (Roche), 4.5 ml sterile distilled water.
4. Once consciousness was lost (usually after five minutes), the mouse was placed face down onto absorbent paper towel and its back was sprayed with absolute ethanol to sterilise this area.
5. A dorsolateral incision was made through the skin and the ovarian fat pad located through this incision.
6. Another, smaller incision was made through the peritoneum directly above the fat pad.
7. The ovarian fat pad was grasped with blunt forceps and gently eased out through the incisions bringing with it the ovary, oviducts and upper part of the uterus.
8. The mouse was then placed onto the stage of a stereo dissecting microscope which was focused on the exposed uterine horn.
9. A hole was made into the uterine horn by insertion of a 25 gauge hypodermic needle (Becton Dickinson), into which the end of the transfer pipette was inserted. The embryos were gently expelled into the uterus from the transfer pipette.
10. After easing the uterus, ovary and fat pad back into the body cavity using blunt forceps, the peritoneum was closed with a single silk suture, and the skin wall by two or three Michel wound clips.
11. The female was kept warm until full consciousness was attained, when she was placed into a clean cage along with other foster mothers.
12. Three days prior to birth, pregnant females were separated into individual cages.

13. A litter of pups resulting from embryo transfer was usually born seventeen to eighteen days following the procedure.

4.10 Analysis of Progeny From Embryo Transfer.

The contribution of the ES cells to the offspring resulting from morula aggregation could be assessed visually by use of pigmentation markers. The chinchilla (c^{ch}) allele carried by the ES cells affects melanocytes which determine coat colour pigmentation. Any contribution to the melanocytes of the offspring from the ES cells will be manifested as hair with reduced pigmentation (yellow pigment affected more than black pigment) when compared to the agouti coat colour contributed by the F_1 C57BL/6 X CBA morulae. The pink eye dilution allele carried the ES cells also manifests itself visually. This allele again affects melanocytes, causing some degree of non-pigmentation (pink) in the eyes, and has the effect of reducing coat pigmentation but affects black pigmentation more than yellow. Therefore, the combined effect of these marker alleles is to allow visual detection of ES cell contribution by loss of pigmentation in the eye and reduced pigmentation of the coat hairs.

CHAPTER 5

HIT AND RUN TARGETING IN ES CELLS WITH ONE STEP ANALYSIS

5.1 INTRODUCTION

Many strategies have been devised to introduce subtle, precise mutations into ES cells. These are described in detail in section 1.2. This thesis describes the use of one of these methods, the 'hit and run' technique (also known as 'in out' targeting), to introduce mutations into *Cftr* in mouse ES cells.

5.1.1 Strategy.

The 'hit and run' technique is a two step procedure requiring two separate homologous recombination events (section 1.2.4.2 and figure 5.1). Initially, a gene targeting experiment is conducted involving large scale clone picking and screening for cells in which the targeting vector has integrated into the target site, the 'hit' event. Following their identification, a negative selection is applied to correctly targeted clones to pick out those cells which have subsequently lost the targeted vector, through the desired intrachromosomal recombination 'run' event. Clones surviving the second round of selection are again picked, grown up and analysed.

These two rounds of selection and clone picking, each requiring analysis, are very labour-intensive and time consuming. Growing up each clone to produce sufficient cells for DNA analysis, and generating frozen stocks, significantly increases the passage number and therefore the opportunity for a deterioration of the condition of the cells, whether through microbial contamination, development of karyotypic abnormalities, or loss of germ line competence.

Through consideration of these aspects, we reasoned that this two step analysis was not strictly necessary. It was argued that only correctly targeted clones will have a duplication of the target DNA and will be capable of losing the integrated vector through intrachromosomal recombination within the duplicated region. Therefore, it would not be necessary to screen the clones surviving the first (positive) selection, which represent targeted and randomly integrated transfected cells. By putting these cells straight onto the second selection, all randomly targeted clones should die as they would still retain the *tk* gene, and only targeted cells which have lost *tk* gene

expression due to excision of the vector, (the 'run' event), should survive. Some cells will inevitably survive Gancyclovir selection through inactivation rather than actual loss of the *tk* gene, but these events should be extremely rare, making these cells a small minority. Therefore, it should be feasible to pick and screen clones only after the second round of selection.

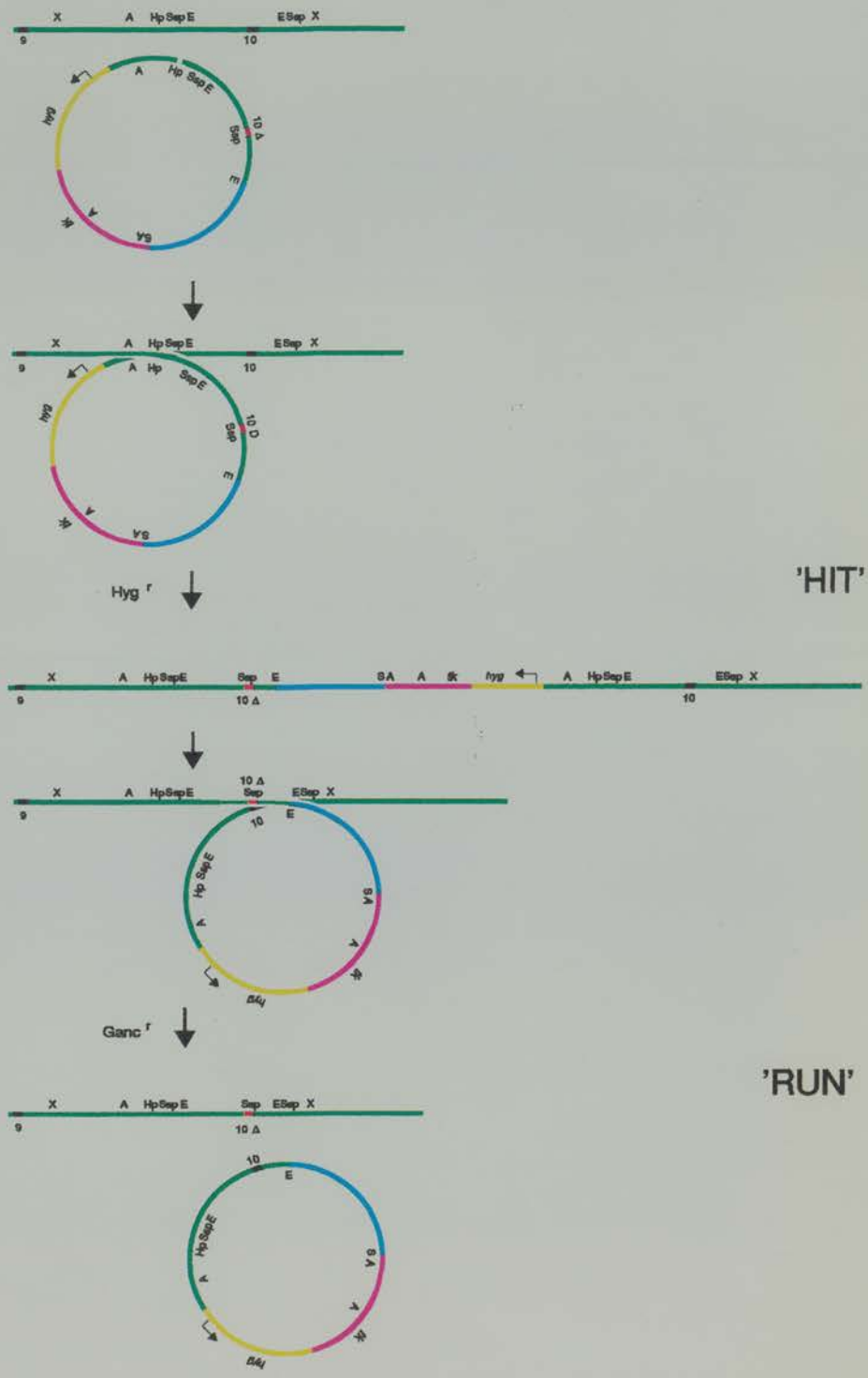


Figure 5.1 'Hit and Run' gene targeting strategy.

5.2 HIT AND RUN TARGETING VECTORS

Vectors employed for the introduction of subtle mutations through the 'hit and run' pathway are a modified version of an insertion vector (section 1.2.4.1). As for insertion vectors, a positive selection gene (e.g. Neomycin or Hygromycin) is necessary for the selection of transfected cells, but for 'hit and run', a negative selection gene (e.g. *tk*) for selection of 'run' events is also required. The vector carries the homologous genomic targeting sequences employed in insertion vectors, but for this purpose, the mutation to be introduced is also present in these targeting sequences.

5.2.1 Vector Construction

The vectors used in this study were constructed by Dr Paul Dickinson, and were based upon the vector pIV3.5H used by Dorin *et al.* (1992) to disrupt *Cftr* in ES cells. This vector was modified as follows (Fig. 5.2):

(I) The Neomycin selection cassette was replaced with a Hygromycin-*hsvtk* fusion gene from the plasmid tgCMV/HgTK (a kind gift from Dr Steven Lupton). This selection cassette produces a fusion protein with the ability to confer both resistance to Hygromycin selection and sensitivity to Gancyclovir selection (Lupton *et al.* 1991). Expressing both positive and negative selection from a single promoter offers the advantage that any gene inactivation/suppression mechanism should act equally upon both selections. This is important when employing a negative selection screen which selects for loss of gene expression. With this fusion gene, any such inactivated cells should not be Hygromycin resistant and therefore would not survive the positive selection round.

(II) The genomic sequences carried by the vector pIV3.5H were extended by the addition of 0.8 kb of genomic sequence, bringing the overall length of genomic homology to 4.3 kb. This additional genomic sequence

included a mutant exon 10, in which a CF-associated mutation had been introduced by use of PCR mediated mutagenesis (Cormack 1994).

Prior to transfection, it is necessary to linearise insertion vectors within the region of homology generating a double strand break which stimulates single, reciprocal recombination with the chromosomal target. Unfortunately, the *Asp 718* restriction site which had been used previously for linearisation of the insertion vector pIV3.5H in high frequency targeting of *Cftr,* was no longer unique, as modification of pIV3.5H had resulted in the introduction of further *Asp 718* restriction sites. Restriction analysis of the genomic sequences was therefore necessary to identify another unique restriction site which was suitable for linearisation of the vector. One such site was identified, an *Hpa* I restriction site, which was situated 1.2 kb from the 5' end of the genomic targeting sequences (Fig. 5.2), and gave a more even distribution of genomic sequences on either side of the double stranded break than the *Asp* 718 site, (1.2 kb 5' and 3.1 kb 3', compared to 0.4 kb 5' and 3.1 kb 3' with *Asp* 718 restriction of vector pIV3.5H).

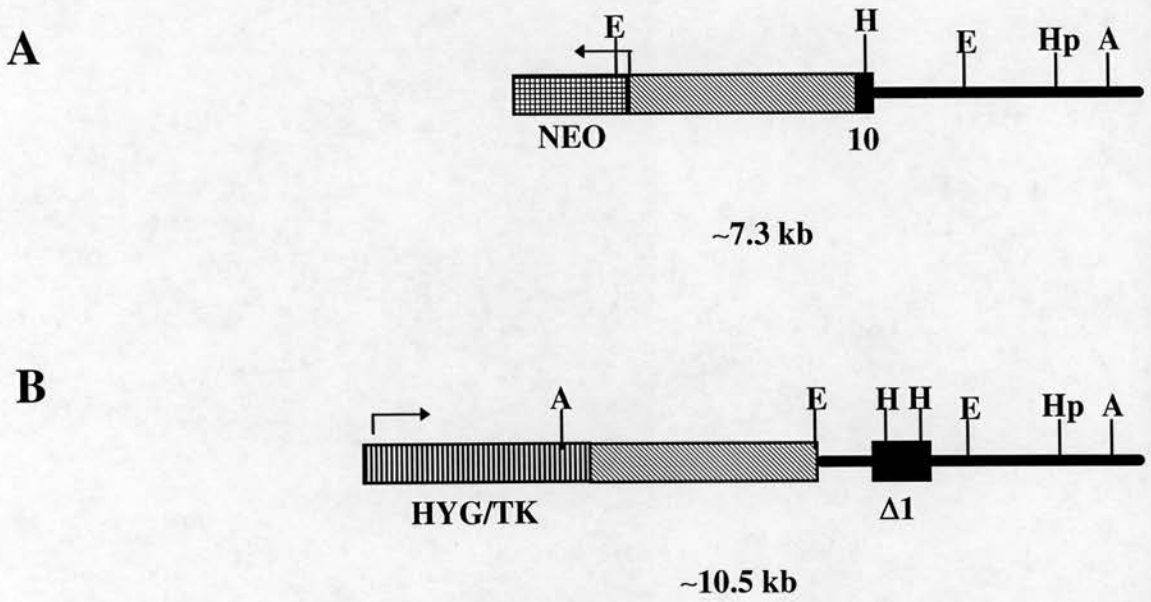


Figure 5.2 *Cftr* targeting vectors.

The solid black line is the region of *Cftr* homology, and the black box represents exon 10. The plasmid sequences are shown as a diagonally hatched region whilst the Neomycin selection gene is indicated by a square hatched area, and the Hygromycin/*hsvtk* fusion gene by a vertically lined area. The arrows indicate the direction of transcription, and Δ the location of a CF-associated mutation. Restriction sites are *Asp 718* (A), *Eco RI* (E), *Hpa I* (Hp), and *Hind III* (H).

(A) Insertion vector pIV3.5H. (B) 'hit and run' vector pHTHRΔ.

5.3 HIT AND RUN TARGETING

5.3.1 Experiment HR1

5.3.1.1 Method

The ES cell line E14 was grown up on a PEF feeder layer (section 2.1), and 10^7 cells of a passage 23 culture were transfected by electroporation with 100 μg of *Hpa* I-linearised vector pHTHRΔF508. Following transfection, the cells were seeded into ten 100 mm gelatinised tissue culture plates (10^6 cells/plate), and incubated overnight. Twenty four hours post-transfection, Hygromycin selection was imposed and maintained until colonies were visible. Eight days after transfection, plates containing Hygromycin resistant colonies were trypsinised and seeded onto PEF feeder layers in fresh plates. Gancyclovir selection was imposed twenty four hours later and maintained until clones were visible. Once visible, the clones were picked and grown up to generate cells for frozen stocks and DNA analysis.

5.3.1.2 Results and analysis.

Clone numbers.

Twenty eight Gancyclovir resistant clones were obtained by this 'hit and run' targeting experiment, and the numbers are detailed in table 5.1. A targeting frequency of 1 in 50 was achieved by Dorin e*t al.* (1992) when targeting this locus with a similar vector. Assuming the same targeting frequency for this experiment, approximately four targeted clones would be expected in the two hundred Hygromycin resistant clones obtained from the first round of selection. With only twenty Hygromycin resistant clones per plate, it would be anticipated that only four out of the ten plates would contain a targeted clone and therefore give clones which survive Gancyclovir selection. As seen below, six plates contained Gancyclovir resistant clones and four did not, which is still within the order of what was expected and implies that at least six independent and correctly targeted clones were obtained after Hygromycin selection. All the Hygromycin resistant clones present on plates 7, 8, 9, and 10 probably represent randomly targeted clones which were not able to excise the plasmid through intrachromosomal recombination, and so consequently died when Gancyclovir was imposed.

Table 5.1 Numbers of clones obtained at each stage of experiment HR1.

PLATE No.	No. HYG^R CLONES	No. GANC^R CLONES
1	AVERAGE OF 20 CLONES PER PLATE	5
2		2
3		13
4		2
5		3
6		3
7		0
8		0
9		0
10		0

Analysis of Gancyclovir resistant clones.

If successful, this experiment should introduce a three base pair deletion, as the only alteration to *Cftr* in the ES cell clones. Detection of such a small change in gene structure requires a sufficiently sensitive technique, and for this experiment it was envisaged that this mutation could be detected by the use of allele-specific PCR (Okayama *et al.* 1989).

Primers were designed which, when used in different combinations, could distinguish between all of the possible clone genotypes (figure 5.3).

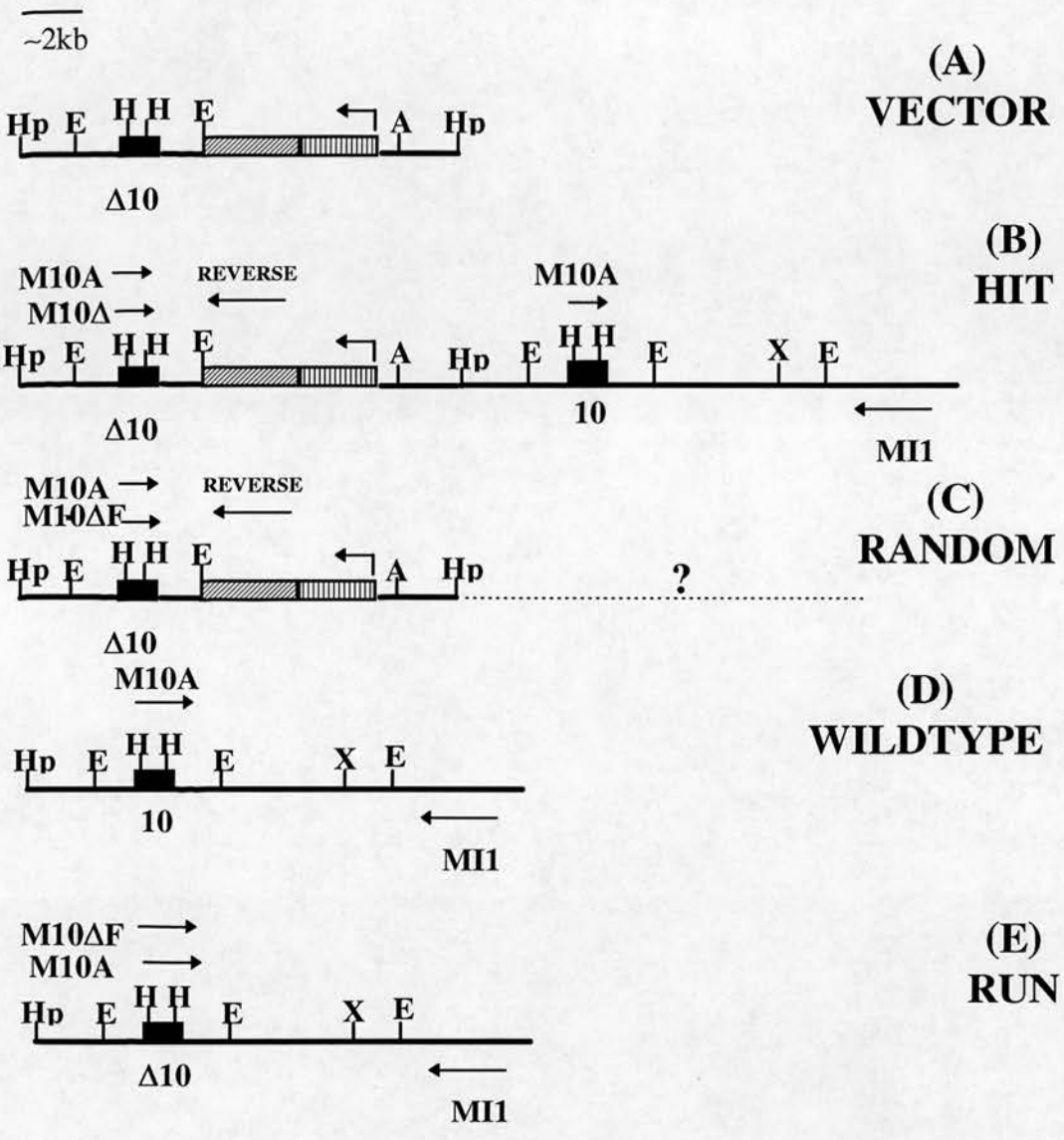


Figure 5.3 The possible genotypes of clones from 'Hit and Run' targeting.
The solid black line represents *Cftr* genomic sequences, the black rectangle exon 10, and the position of a CF-associated mutation is marked by Δ. Unknown genomic sequences are indicated by a dotted line. Plasmid sequences are shown as a diagonally hatched region, and the Hyg/TK fusion gene by vertical hatching. The direction of transcription is indicated by a bent arrow. Restriction sites are *Asp 718* (A), *Eco RI* (E), *Hpa I* (Hp), *Hind III* (H), and *Xba I* (X). The annealing sites of PCR oligonucleotides and the direction of priming are indicated by arrows. (A) The structure of pHTHRΔ 'Hit and Run' targeting vector. (B) The predicted mutant genomic structure around exon 10 of *Cftr* following homologous recombination with vector pHTHRΔ. A 'hit' event. (C) Random, non-homologous integration of vector pHTHRΔ with unknown flanking genomic sequences. (D) Genomic structure around exon 10 of wildtype *Cftr* allele. (E) The desired mutant allele of a 'run' event, differing from the wildtype allele only by the presence of a mutation (Δ) in exon 10.

A PCR reaction was conducted on HR1 clone DNA, utilising the primer M10ΔF which is specific for the ΔF508 mutation, and primer MI10 which primes from a 3' region external to the targeting region. This combination is therefore diagnostic for a 'run' event which has retained the ΔF508 mutation. No PCR products were obtained from this reaction, but as there was no positive control template DNA available, this in itself is inconclusive. The possibility that the absence of product is due to failure of the PCR reaction rather than absence of the correct template cannot be ruled out.

A second PCR reaction was conducted, this time employing the primers M10AI and MI10 which amplify across exon 10 out into genomic DNA flanking the 3' side of the vector. This reaction would amplify DNA from a 'run' event which has retained or lost the mutation, and the 3' duplicated region (predicted to be wildtype) from a 'hit' event, but should not amplify DNA from a randomly targeted clone. The wildtype allele on the non-targeted chromosome will also be amplified in all genotypes. PCR products of the correct size were obtained for all clones and were run out on a 2% agarose gel, blotted, and then analysed by allele specific hybridisation (section 3.6). The filter containing the PCR products was probed with the oligonucleotide M10ΔF, which should hybridise to amplified exon 10 DNA containing the ΔF508 mutation. This probe however, exhibited some non-specific hybridisation as shown by the signal obtained in the lane containing non-transfected E14 DNA (figure 5.4). After employing the stringent washing conditions of 2 X SSC for fifteen minutes at 75°C, this aberrant signal was still visible in the E14 lane, but was significantly stronger in the positive control lane containing amplified vector DNA bearing the ΔF508 mutation (PCR products from M10ΔF and reverse primers). After accounting for the intensity of Ethidium bromide staining, the signal in the lanes of HR1 clones was similar to that of the E14 negative control and less intense than the positive control. Despite the non-specific hybridisation exhibited, this result suggests that the ΔF508 mutation was not present in the products of this PCR reaction and that the amplified DNA represents a wildtype allele. This could mean that the clones are all from a 'run' event which has reverted to wildtype (or wildtype cells which have come through the selections), but as at least one wildtype allele will be amplified in all genotypes, does

not exclude the possibility that the clones still contain the vector. The mutant exon 10 of an integrated vector would not be amplified in this PCR reaction.

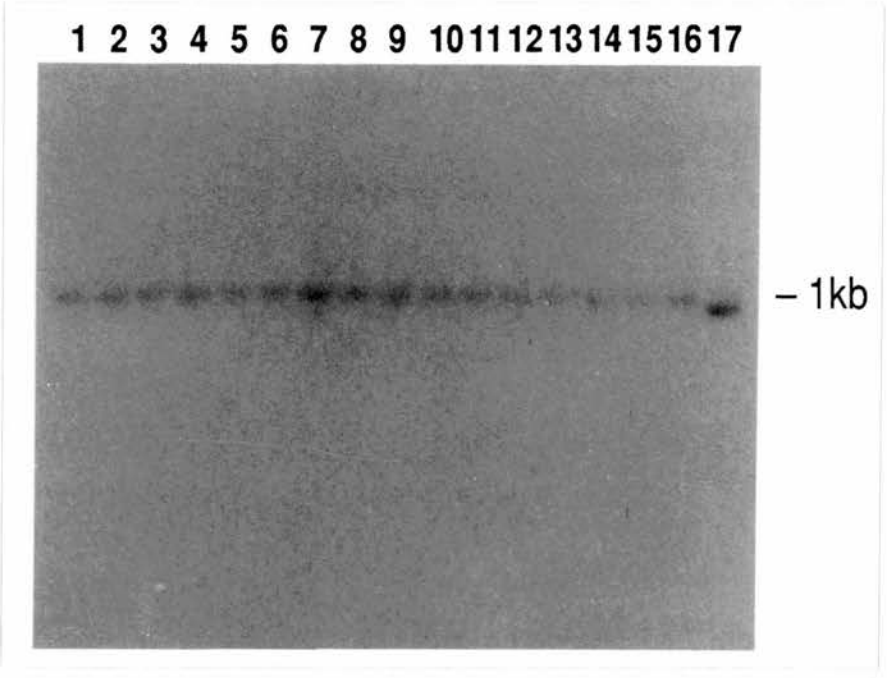


Figure 5.4 Hybridisation of ΔF508 allele-specific oligonucleotide (M10ΔF) to PCR-amplified *Cftr* exon 10 DNA of HR1 clones.

Genomic DNA was amplified using primers M10AI and MI10, blotted and probed with the ΔF508 allele-specific oligonucleotide M10ΔF.
Lane 1 is a negative control, non-transfected E14 DNA, and lanes 2-16 HR1 clones. Lane 17 is a positive control of pHTHRΔF508 vector DNA amplified with primers M10ΔF and reverse primers generating a ΔF508 mutation containing product.

In order to investigate whether the targeting vector was still present in these clones, a PCR reaction was carried out using the primers M10ΔF and Reverse. These amplify from the ΔF508 mutation across into plasmid sequences carried by the vector, and so are specific for integrated vector DNA but does not distinguish between correctly targeted and randomly targeted vector DNA. Products of the expected size were generated by this PCR reaction (figure 5.5) suggesting that both the mutation and the vector are present in the clones. A faint band can be seen in the negative control which was set up in the absence of a DNA template, and probably represents contamination of the reaction with vector DNA. However, the bands given by the HR1 clones (apart from lanes 6 and 7), are much stronger than this contaminant band and are unlikely to have arisen through vector contamination.

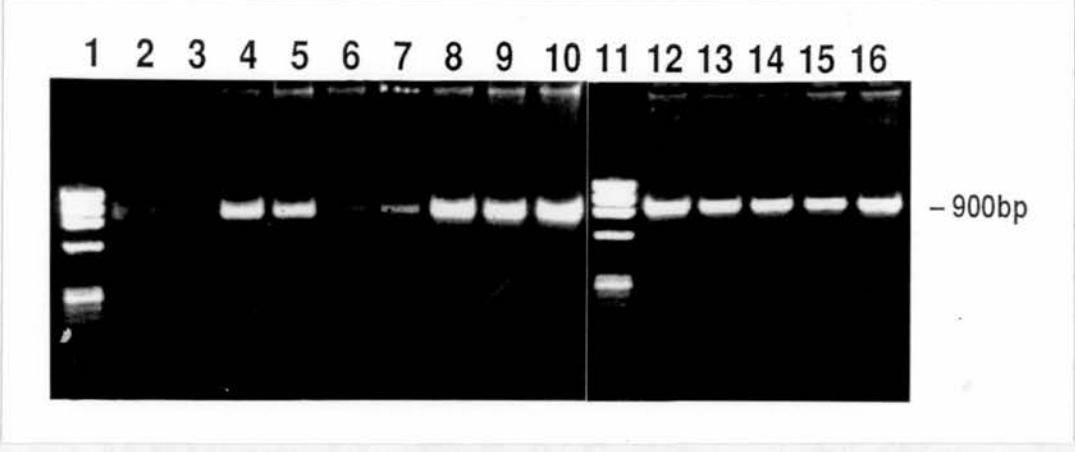


Figure 5.5 Amplification of vector sequences in HR1 clone DNA.
An Ethidium bromide stained 1% agarose gel showing HR1 clone DNA amplified in a PCR reaction with primers M10ΔF and reverse primers.
Lanes 1 and 11 are PhiX174 *Hae III* size markers. Lanes 2 and 3 are negative controls of amplification without DNA template, or with a non-transfected DNA template respectively. Amplified DNA from HR1 clones is present in lanes 4-10 and 12-15. Lane 16 is the positive control containing amplified DNA from a cell line known to have integrated a similar ΔF508 targeting vector (obtained from Dr Julia Dorin).

5.3.1.3 Conclusions

This 'straight through' 'hit and run' strategy seems to be at least partly successful as Gancyclovir resistant clones were obtained, and appear to be a result of initial transfection with the targeting vector. However, it would seem that all of the clones obtained in this experiment were 'false positives' in that they had survived Gancyclovir selection despite containing the vector. This is probably due to inactivation of the *tk* gene in these clones which could be caused by point mutation or the influence of flanking DNA in randomly integrated clones. The desired excision of the vector following integration in the 'hit' stage had not occurred.

Analysis of these clones was not straightforward and a much simpler and reliable method would be needed, especially in view of the large number of transgenic mice which would be generated and require analysis if this targeting was successful. Therefore it was decided to make use of codon 'wobble' and alter the nucleotide sequence of the vector for future experiments so that a novel restriction site is created by the mutation, whilst the amino acid sequence remains unchanged.

5.3.2 Experiment HR2

Another 'straight through' 'hit and run' experiment was conducted along the same lines as experiment HR1. This time a targeting vector was used bearing another CF associated mutation, the ΔI506/7 mutation, which creates an *Ssp I* restriction site to facilitate clone analysis. It had been decided at this point that 'hit and run' targeting for this thesis should concentrate upon introducing the ΔI506/7 mutation into ES cells, and that targeting of the ΔF508 mutation would be carried out in parallel by others in our group. This was due to the desirability of a mouse model carrying the most common CF mutation, the ΔF508 mutation, and the pressure to produce this mouse model quickly. The characterisation of a ΔF508 mouse would need to be conducted speedily and would not allow time for the acquisition of the necessary level of skill in the appropriate techniques. However, although a ΔI506/7 mouse would be of considerable and complimentary interest, its creation would not attract the same degree of urgency and would therefore be better suited to a project for a PhD thesis.

The ΔI506/7 mutation represents a deletion of an Isoleucine amino acid at either codon 506 or 507 (both are Isoleucines., Kerem *et al.* 1990). This mutation accounts for approximately 0.2% of CF mutations, making it around the fourteenth most common CF-causing mutation (The Cystic Fibrosis Genetic Analysis Consortium 1994). The proximity of this single amino acid deletion at codon 506 or 507, to the major CF-associated mutation ΔF508 (also a single amino acid deletion) at codon 508, makes a comparison between the two mutations worthwhile. In common with ΔF508 CFTR, the mutant protein arising from the ΔI506/7 mutation has also been shown to be subject to biosynthetic arrest, becoming mislocalised within the cell (Cheng *et al.* 1990). Due to its low frequency, very few ΔI506/7 homozygous patients have been reported, however this mutation has been designated a 'severe' allele based upon the observation that it confers a pancreatic insufficient phenotype in ΔI506/7/ΔF508 compound heterozygotes (Kerem *et al.* 1990., Kristidis *et al.* 1992).

The ΔI506/7 mutation was engineered into the targeting vector in such a way that it created a diagnostic restriction site (figure 5.6). This was achieved by making use of codon 'wobble', a phenomenon which allows multiple codons to specify the same amino acid due to a requirement for accurate base pairing in the first two bases of a codon only.

A Wildtype sequences

Nucleotide sequence	5'	**AAT**	**ATC**	**ATC**	**TTT**	**GGT**	3'
Amino acid sequence	N	**N**	**I**	**I**	**F**	**G**	C

B Mutant ΔI506/7 sequences

Nucleotide sequence	5'	**AAT**		**ATC**	**TTT**	**GGT**	3'
Amino acid sequence	N	**N**		**I**	**F**	**G**	C

C *Ssp I* restriction enzyme recognition sequence

5' **AATATT** 3'

D Vector sequences ↓

Nucleotide sequence	5'	**AAT**		**ATT**	**TTT**	**GGT**	3'
Amino acid sequence	N	**N**		**I**	**F**	**G**	C

Figure 5.6 Alteration of vector ΔI506/7 sequences to introduce a diagnostic restriction site.

(A) Represents wildtype nucleotide and amino acid sequences at codons 505 to 509, and (B) the mutant sequence of the ΔI506/7 mutation in which an Isoleucine has been deleted at codon 506. (C) Shows the recognition sequence of the restriction enzyme *Ssp I,* and (D) the altered mutant nucleotide sequence which creates the *Ssp I* site whilst leaving the amino acid sequence unchanged. The nucleotide bases are represented as (A) adenine, (T) Thymine, (G) Guanine, and (C) Cytosine. The amino acids are denoted as (N) Asparagine, (I) Isoleucine, (F) Phenylalanine, and (G) Glycine. The site of cutting within this sequence is indicated by the arrow.

5.3.2.1 Method

A passage nineteen culture of E14 ES cells were grown up on a PEF feeder layer, and 10^7 cells transfected with 100 µg of *Hpa I*-linearised vector pHTHRΔI. The cells were plated out as before (see HR1) and Hygromycin selection imposed after twenty four hours.

It was thought that in the previous experiment, the time between Hygromycin and Gancyclovir selection, (twenty four hours), may not have been sufficient to allow excision of the vector (the 'run' event), to occur. In the original paper by Hasty *et al.* (1991) the highest rate of vector excision achieved was 4 X 10^{-3} cells. Some of the small Hygromycin resistant clones may not contain as many as 10^3 cells and occurrence of this excision event would therefore be unlikely. To determine if this has some effect on the number of Gancyclovir clones obtained, five of the plates containing Hygromycin resistant clones (series A), were trypsinised into fresh gelatin coated plates seven days after Hygromycin selection and maintained without selection for two days to allow cell numbers to increase before Gancyclovir selection was imposed. The other five plates (series B), were maintained on Hygromycin selection for ten days, and then trypsinised into fresh gelatinised plates where Gancyclovir selection was imposed immediately. For this experiment, Gancyclovir selection was carried out in the absence of PEF feeder layers to eliminate the possibility of their presence playing some part in allowing cells containing the vector to survive Gancyclovir selection. Ten days after Gancyclovir selection was imposed, resistant clones were picked and seeded into a 96 well plate. Unfortunately, the medium used to feed these cells must have been contaminated and after an overnight incubation, all clones were contaminated and discarded.

5.3.2.2 Results

Despite losing all of the clones through contamination, some useful data was obtained from this experiment concerning the numbers of clones surviving each selection (table 5.2).

Apart from plate A3, there seems to be no significant difference between the number of Gancyclovir resistant clones obtained in series A than with series B plates. The two days of selection-free culture of series A allowed the cell numbers to increase before Gancyclovir selection was imposed, and should have increased the opportunity for a 'run' event to occur. The large number of Gancyclovir resistant clones present on plate A3 may be significant. Perhaps these clones represent the number of cells which arise from one or more true 'run' events, whereas the smaller number of clones present on other plates represent a small number of cells which are able to survive Gancyclovir selection through other, less frequent, means of loss of gene expression.

Table 5.2. Numbers of clones obtained at each stage of experiment HR2.

PLATE No.	No. HYG^R CLONES	No. $GANC^R$ CLONES
A1	12	1
A2	8	0
A3	17	>100
A4	14	0
A5	17	5
B1	33	0
B2	18	0
B3	24	0
B4	21	6
B5	18	12

5.3.2.3 Conclusions

It is difficult to draw any conclusions from this experiment as none of the clones was available for DNA analysis. The numbers of Gancyclovir resistant clones obtained was promising, and it would have been interesting to establish if the large numbers of clones on plate A3 had survived Gancyclovir selection through a different mechanism to those clones present on other plates.

5.4 DISCUSSION

From these experiments it was difficult to establish whether this targeting strategy was successful. Although this 'straight through' strategy is time and labour saving, it had the disadvantage that only the end result was analysed and no information was obtained regarding events preceding this last step. It could be argued that *Cftr* is not being targeted at all during the first step, and that all of the Hygromycin resistant clones represent random targeting events which are only able to survive Gancyclovir selection by inactivation of the *tk* gene, rather than through excision of the vector. The more time consuming method of picking and analysing the Hygromycin resistant clones after the transfection would provide information on the targeting efficiency at the *Cftr* locus, and the 'run' step could then be carried out only on those correctly targeted clones.

CHAPTER 6

TARGETING THE *Cftr* LOCUS

6.1 INTRODUCTION

On the basis of previous experiments described in chapter 5, it was decided to conduct the 'hit and run' strategy in two separate stages, analysing the clones resulting from each step. This should provide information not available from previous experiments as to whether *Cftr* is being correctly targeted, the manner in which the vector is integrating, and the frequency at which this targeting occurs. It would also enable targeted clones to be thoroughly examined before proceeding onto the second stage, and therefore maximises the chance of achieving the desired result.

The first step in this procedure involved targeting the *Cftr* locus with the 'hit and run' vector, pHTHRΔI (described in section 5.2). In contrast to previous experiments, the clones resulting from this initial step were picked individually and screened for those in which the vector had integrated through homologous recombination into the target site.

6.2 HR3 TARGETING EXPERIMENT

6.2.1 Method.

A passage 19 culture of E14 ES cells was grown up and 10^7 cells transfected by electroporation with 100 μg of the 'hit and run' targeting vector pHTHRΔI, as described in section 2.5. Hygromycin selection was carried out as before, and once visible, clones were picked and grown up for frozen stocks and analysis.

6.2.2 Results.

6.2.2.1 Clone numbers.

The numbers of Hygromycin resistant clones obtained from this transfection are shown in table 6.1. The transfection efficiency of this experiment was 1×10^{-3}. No clones were obtained from the non-transfected control plates, indicating that the selection was working efficiently. A total of ninety six clones were picked and grown up, sixty six of which were taken through to DNA analysis.

Table 6.1 Numbers of Hygromycin resistant clones obtained in HR3 targeting experiment.

PLATE No.	TOTAL No. HYG^R CLONES	No. CLONES ANALYSED
CONTROLS	0	0
1	4	4
2	17	12
3	14	14
4	8	0
5	14	11
6	4	4
7	8	8
8	10	7
9	8	3
10	25	3
TOTAL	112	66

6.2.2.2 Screening for 'Hit' clones.

Hygromycin resistant clones from experiment HR3 were screened by Southern blot analysis to identify clones in which the vector had integrated into the target site. Genomic DNA was extracted from the clones, subjected to double digestion with the enzymes *Xba I* and *Sal I*, and hybridised to the 0.6EH genomic probe. This probe consists of a 0.6 kb *Eco RI-Hind III* genomic fragment from a region of *Cftr* which lies 3' to the target site and which would be external to a correctly integrated vector (figure 6.1). An *Xba I* fragment which encompasses the target site hybridises with the 0.6EH probe to give a band of approximately 6 kb for the wildtype allele. Homologous recombination of the targeting vector into this target site would increase the size of this *Xba* I fragment by that of the targeting vector (approximately 10.5 kb) to 16.5 kb. The integrated vector introduces a *Sal I* site into this *Xba I* fragment, and correctly targeted clones will therefore generate two fragments, one of which hybridises to the 0.6EH probe to give a band of 8 kb. In this assay, homologous recombinants are identified by presence of a 8 kb mutant band appearing above the 6 kb wildtype band, as shown in figure 6.2.

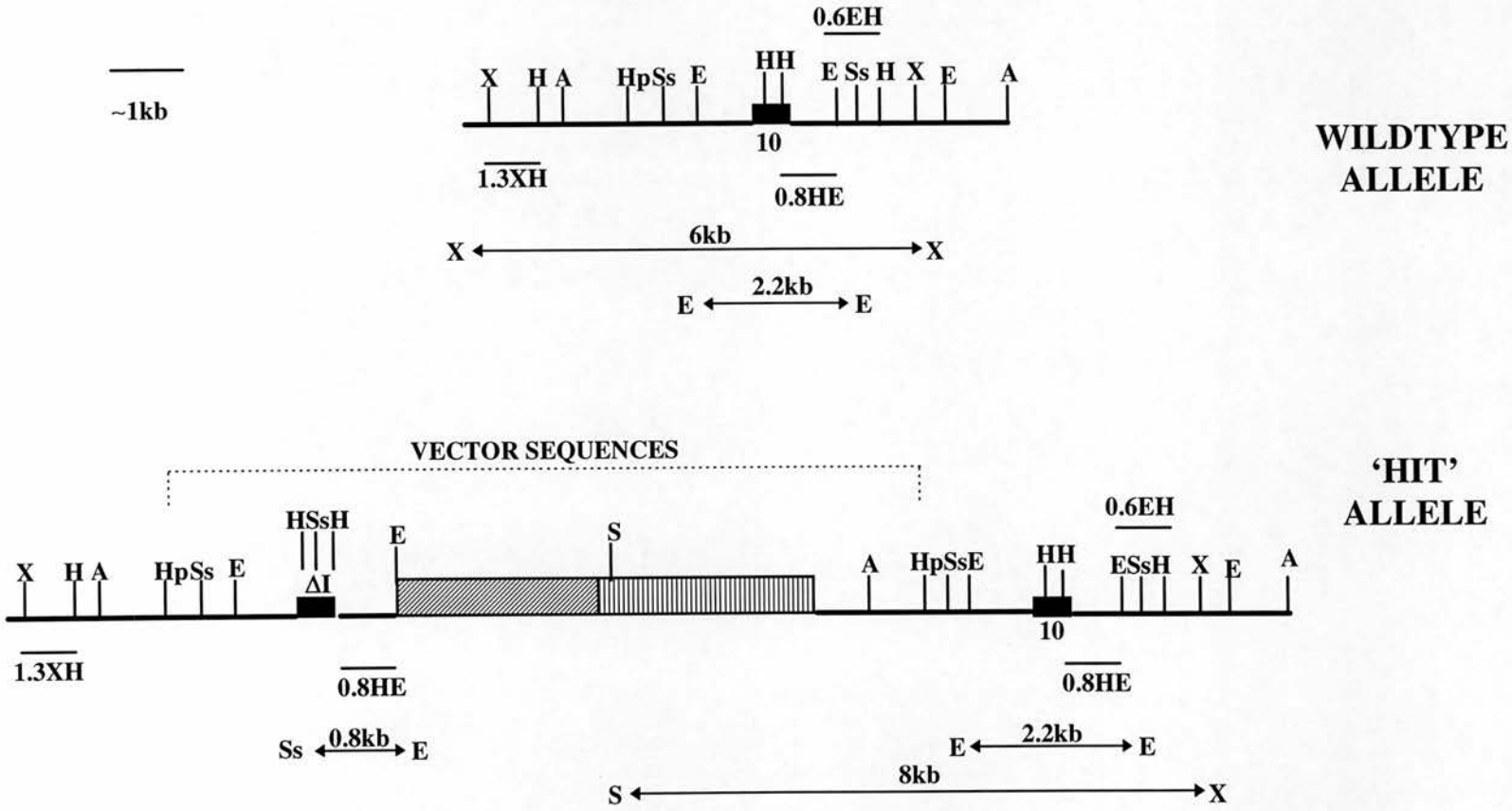


Figure 6.1 Restriction map of exon 10 target site.

The solid black line represents the *Cftr* genomic sequences, the black rectangle exon 10, and the position of the ΔI506/7 mutation is marked by ΔI. Plasmid sequences are shown as a diagonal hatched region, and the Hygromycin-*tk* fusion gene by vertical hatching. The restriction sites are *Asp 718* (A), *Eco RI* (E), *Hind III* (H), *Hpa I* (Hp), *Sal I* (S), *Ssp I* (Ss), and *Xba I* (X). The site of hybridisation of each probe, and the diagnostic fragments which they detect are indicated beneath each allele.

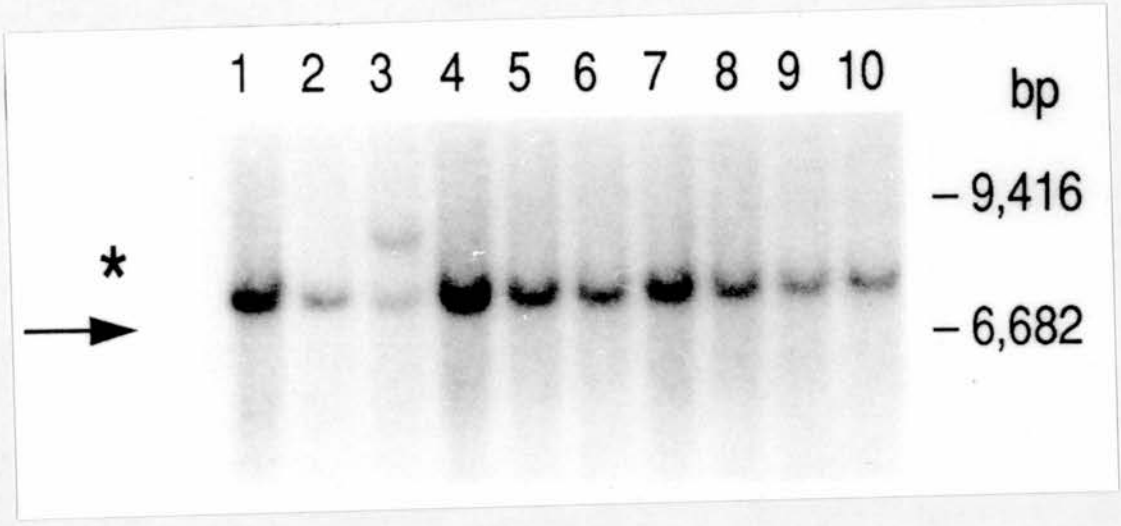


Figure 6.2 Screening for targeted 'hit' clones.

Southern blot analysis of HR3 clone DNA *Xba I-Sal I* digests probed with the 0.6EH probe. Lane one is the non-transfected control and displays the 6 kb band of the wildtype allele, indicated by the arrow. Lanes 2 to 10 contain HR3 clone DNA. The presence of the upper 8 kb band in lane three (denoted by *) in addition to the 6 kb wildtype band, is indicative of a correct targeting event. The position of the closest lambda *Hind III* marker bands are indicated to the right of the blot.

Screening of HR3 clones in this manner detected three correctly targeted 'hit' clones designated 3.7, 3.10, and 7.8 (figure 6.3). A total of 66 Hygromycin resistant clones were analysed, giving a targeting frequency of 1 homologous recombinant in every 22 transfected clones for this experiment (4.6%). The equal dosage of bands indicates the expected ratio of one wildtype to one mutant allele, but does not exclude the possibility of multiple copy integration, as this probe would only pick up the most 3' copy of multiple, tandemly integrated vectors.

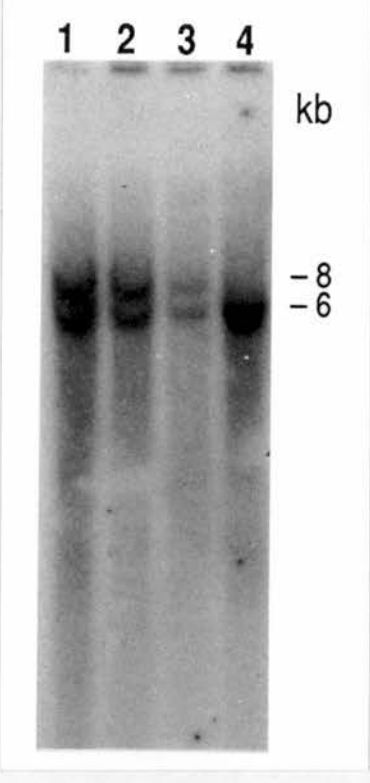


Figure 6.3 Southern blot of three HR3 targeted 'hit' clones.

This figure shows the Southern blot result of the three targeted HR3 clones identified by probing *Xba I-Sal I* digests with the 0.6EH probe. Lanes 1 to 3 are the targeted clones which exhibit the 8 kb band of the targeted allele in addition to the 6 kb band of the wildtype allele. Lane 4 is the non-transfected control displaying only the 6 kb wildtype band.

6.2.2.3 Confirmation of 'Hit' clones.

Confirmation that these 'hit' clones were the result of the predicted integration of a single copy of the targeting vector into the targeted site was obtained by further Southern blot analysis.

The filter of HR3 'hits' digested with *Xba I* and *Sal I* (figure 6.3) was reprobed with the 1.3XH probe. This is a 1.3 kb *Xba I-Hind III* genomic fragment, this time from a region of *Cftr* 5' to the target site which is predicted to be external to a correctly targeted vector (figure 6.1). This hybridised to a band of the predicted size of 6 kb for the wildtype allele, and 8.5 kb for the mutant allele, which represents the second of the fragments generated by the unique *Sal I* site introduced by the vector into the *Xba I* fragment encompassing the target site (figure 6.4). This confirms correct targeting and the integrity of the 5' region of the target site.

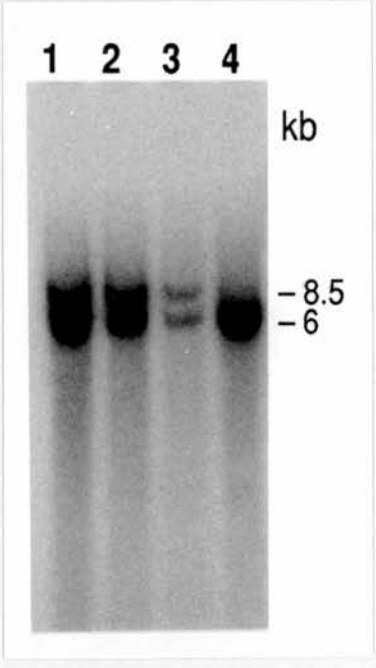


Figure 6.4 Confirmation of HR3 'hit' clones.
The Southern blot of HR3 *Xba I-Sal I* digests reprobed with the 1.3XH probe. As before, lanes 1 to 3 are the targeted clones which display the 8.5 kb band of the mutant allele, in addition to the 6 kb band of the wildtype allele. Lane 4 is the non-transfected control exhibiting only the 6 kb wildtype band.

6.2.2.4 Screening of 'Hit' clones for presence of mutation.

The presence of the ΔI506/7 mutation in these 'hit' clones was ascertained by probing *Ssp I-Eco RI* genomic digests with the 0.8HE probe. This probe is a 0.8 kb *Hind III-Eco RI* genomic fragment which is present in the genomic targeting sequences carried by the vector (figure 6.1). Hybridisation with this probe detected the expected 2.2 kb fragment of both the wildtype allele of the non-targeted chromosome, and the non-mutant, endogenous allele of the targeted chromosome, as well as the smaller 0.8 kb band created by cutting at the unique *Ssp I* site produced by the mutation (figure 6.5).

Screening of the clones by this method indicated that one of the clones, 7.8 had lost the mutation, whilst clones 3.7 and 3.10 had retained the mutation. The dosage of the mutant band versus the wildtype bands suggests that only a single copy of the vector has integrated. The wildtype band is expected to be twice as intense as the mutant band, representing the two wildtype alleles versus the one mutant allele. In this case the mutant allele is actually less than half the intensity of the wildtype, but as this is a small fragment (approximately 800 bp), it may be explained by inefficient DNA transfer. The absence of any unexpected bands on this blot confirms that there are no other copies of the vector integrated elsewhere in the genome of these clones, as the probe was internal to the targeting vector and would have visualised any other copies of the vector.

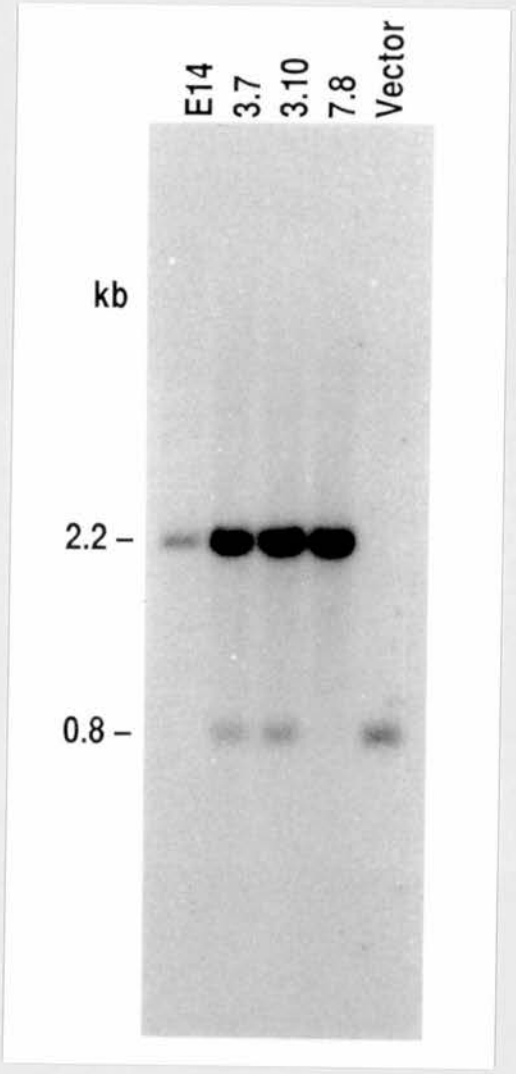


Figure 6.5 Screening HR3 'hit' clones for the presence of the ΔI506/7 mutation. The lane labelled E14 contains DNA from the non-transfected control, and exhibits the 2.2 kb band of the wildtype allele alone. The following 3 lanes contain DNA from HR3 'hit' clones, with the name of each clone indicated above the appropriate lane. These all display the 2.2 kb band generated by both the wildtype allele and the non-mutant exon 10 of the targeted allele. In addition, clones 3.7 and 3.10 exhibit the 0.8 kb band generated by cutting at the novel restriction site introduced by the ΔI506/7 mutation. Vector DNA alone is present in the lane labelled vector, and confirms that the 0.8 kb band is indeed generated from vector sequences.

6.2.2.5 Karyotype analysis of targeted clones.

The karyotype of the three 'hit' clones was checked to ensure that further experiments were only conducted with clones of a good karyotype which should be capable of producing germline chimaeras. Chromosome counts (section 2.1.6) revealed that clones 3.7 and 3.10 had only 2/10 and 1/10 mitotic spreads with the correct chromosome number respectively. Clone 7.8 was slightly better, with 4/10 spreads with the correct chromosomal number, which might be good enough to warrant continuing into the next phase.

6.2.3 Conclusions.

This experiment was successful in targeting *Cftr* at a high frequency of 4.6%, generating three targeted 'hit' clones. Two of these 'hit' clones, 3.7 and 3.10, arose on the same plate (plate three), and so it is possible that they are clonal, both arising from the same targeting event. This would decrease the estimated targeting efficiency to 3%, which would still be considered a high targeting frequency for this locus. However, a greater number of clones originating from plate three were analysed than from any other plate. Therefore, the multiple targeting events identified on this plate are probably a reflection of this. All three of these clones appear to be the result of insertion of the targeting vector into the target site in the predicted manner. The ΔI506/7 mutation was retained in clones 3.7 and 3.10, but the average chromosome number of cells from these clones was too abnormal to warrant continuing with these clones. In contrast, clone 7.8 had a karyotype which was just passable, but this clone had lost the ΔI506/7 mutation. Naturally, only 'hit' clones which have retained the mutation can be used for the next stage, where excision of the vector along with the endogenous wildtype copy generates cells in which the mutation is retained as the only alteration to Cftr. Conducting a 'run' experiment with clone 7.8 would mean that the only genotype possible on excision of the vector would be a reversion to wildtype. A good karyotype is required in order to generate germ line chimaeras from such mutant cells.

In summary, although this experiment was successful in targeting *Cftr*, none of the 'hit' clones generated were suitable for taking forward into the second, 'run' step.

6.3 HR5 TARGETING EXPERIMENT

The targeting experiment HR3 generated three 'hit' clones, none of which were suitable for conducting the 'run' step due to loss of mutation and poor karyotypes. This targeting was repeated again, but in view of the poor karyotypes obtained previously, a new ES cell line (CGR8) which had a stable karyotype and good record of germ line competency was used.

6.3.1 Method.

A passage thirteen culture of CGR8 ES cells was grown up on a PEF feeder layer, and 8 X 10^6 cells were transfected as before (section 2.1.8). Hygromycin resistant clones were selected, picked and prepared for analysis as before.

6.3.2 Results.

6.3.2.1 Clone numbers.

The number of Hygromycin resistant clones obtained is shown in table 6.2. The transfection efficiency of this experiment, 2.8 X 10^{-3}, is slightly higher than that obtained with HR3. This is known to be dependent upon many factors which vary between experiments, and might be a reflection of the different cell lines used. A greater number of clones were obtained overall in this experiment and more were taken through to the analysis stage than in HR3.

Table 6.2 Numbers of Hygromycin resistant clones obtained in HR5 targeting experiment.

PLATE No.	TOTAL No. HYG^R CLONES	No. CLONES ANALYSED
CONTROLS	0	0
1	20	6
2	18	5
3	15	7
4	20	10
5	23	13
6	48	37
7	37	16
8	40	16
TOTAL	221	110

6.3.2.2 Screening for 'hit' clones.

The Hygromycin resistant clones from experiment HR5 were screened by Southern blot analysis as before, by probing *Xba I-Sal I* double genomic digests with the 0.6EH probe (6.2.2.2). This resulted in the identification of fifteen 'hit' clones (figures 6.6 and table 6.3), giving a targeting frequency of 13.6% which is higher than that of the previous experiment. However, one of the 'hit' clones represented by lane 3 of figure 6.6, gave an additional band along with those expected for a targeted 'hit' clone. This band was absent in a subsequent Southern blot utilising a probe internal to the targeting vector (figure 6.8), and therefore cannot be attributed to additional copies of the targeting vector. As this band hybridises to the 0.6EH genomic probe, it might have arisen through a rearrangement of the target site and therefore was not used for future experiments. All other 'hit' clones gave bands of the expected size which confirmed the integrity of the 3' genomic sequences. The equal intensity of these bands confirmed the expected dosage of one mutant to one wildtype allele.

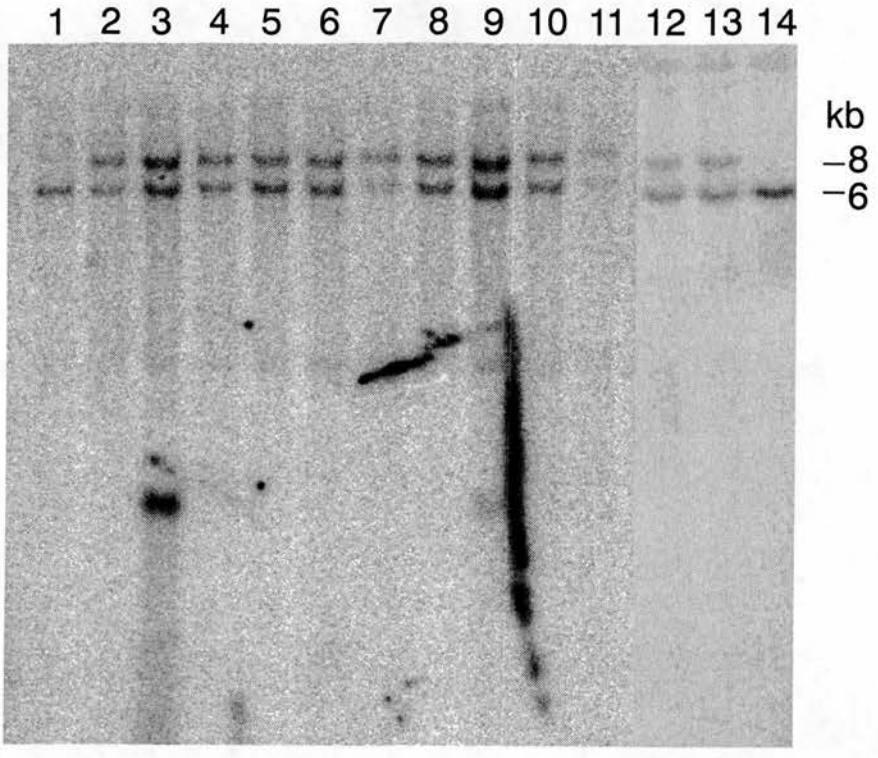


Figure 6.6 Southern blot of HR5 targeted 'hit' clones.
Southern blot showing the bands displayed by the targeted 'hit' clones identified by screening HR5 clones by probing *Xba I-Sal I* digests with the 0.6EH probe. Lane 1 is the non-transfected control which exhibits the 6 kb of the wildtype allele alone. Lanes 2 to 13 are the targeted 'hit' clones identified by the presence of the 8 kb band in addition to the wildtype band. Lane 14 is a randomly targeted clone which displays only the wildtype band with this target site-specific probe.

6.3.2.3 Confirmation of 'hit' clones.

The structure of the 5' region of the targeted site in 'hit' clones was confirmed as in HR3 (6.2.2.3) by reprobing the filter of *Xba I-Sal I* digests (figure 6.6) with the 5' 1.3XH probe (figure 6.7). Bands of the expected size of 6 kb for the wild type allele, and 8.5 kb for the mutant allele were obtained, confirming the integrity of the 5' site.

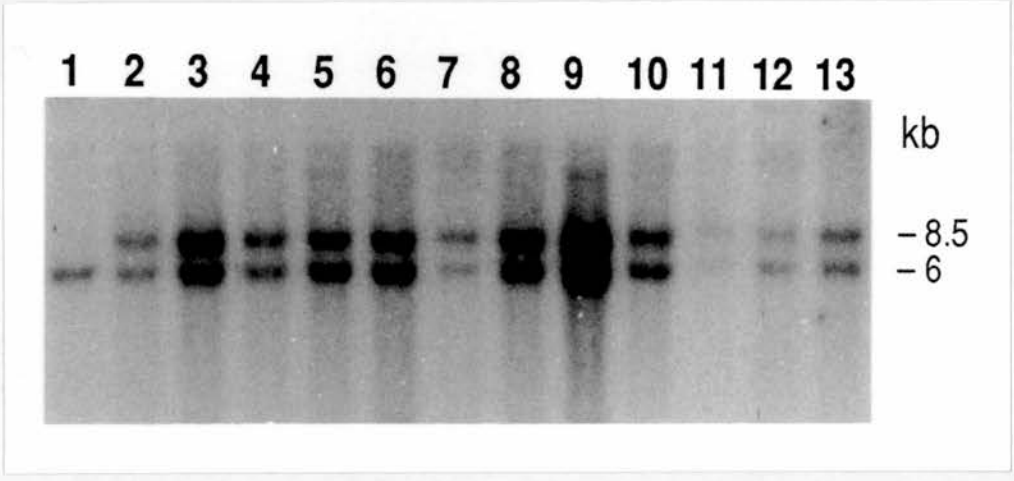


Figure 6.7 Confirmation of targeted clones by reprobing with a 5' probe.
The Southern blot displayed in figure 6.6 was reprobed with the 1.3XH 5' probe. As before, Lane 1 is the non-transfected control which exhibits the 6 kb of the wildtype allele alone. Lanes 2 to 13 are the targeted 'hit' clones identified by the presence of the 8.5 kb band in addition to the wildtype band.

6.3.2.4 Screening for presence of the mutation.

The presence of the mutation was ascertained as before (6.2.2.4) by probing a *Ssp I-Eco RI* Southern blot with the 0.8HE probe which hybridises to genomic sequences carried by the vector (figure 6.8). This identified 12 (80%) 'hit' clones which had retained the mutation, and 3 (20%) clones which lost the mutation (detailed in table 6.3). The absence of any other bands hybridising to this internal probe indicates that there are no other copies of the vector integrated elsewhere in the genome.

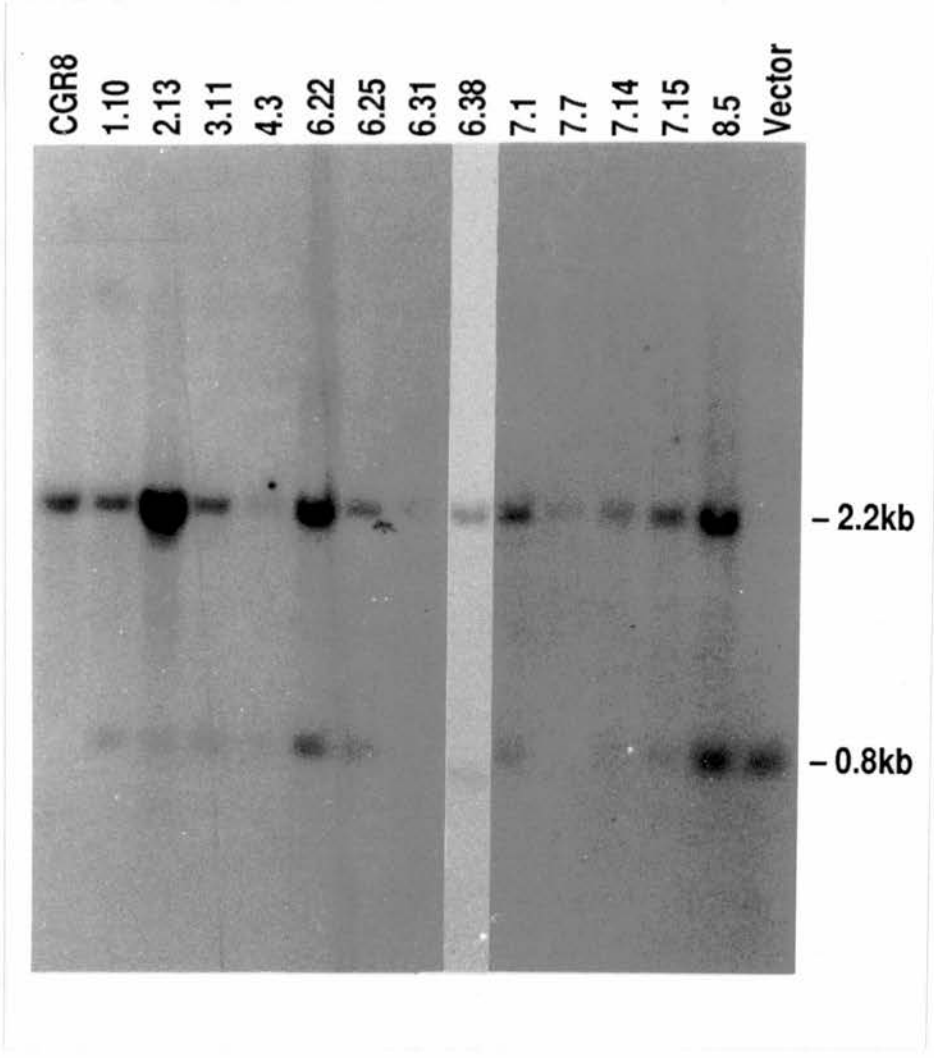


Figure 6.8 Screening HR5 'hit' clones for the presence of the ΔI506/7 mutation. The lane labelled CGR8 contains DNA from the non-transfected control, and exhibits the 2.2 kb band of the wildtype allele alone. The following 12 lanes contain DNA from HR5 'hit' clones, with the name of each clone indicated above the appropriate lane. These all display the 2.2 kb band generated by both the wildtype allele and the non-mutant exon 10 of the targeted allele. In addition, some clones exhibit the 0.8 kb band generated by cutting at the novel restriction site introduced by the ΔI506/7 mutation. The lane labelled 8.5 is a non-targeted clone. Vector DNA alone is present in the lane labelled vector, and confirms that the 0.8 kb band is indeed generated from vector sequences.

6.3.2.5 Chromosome analysis of HR5 'hit' clones.

The chromosome count of some of the 'hit' clones was analysed and found to be consistently good, with between 90-100% of mitotic spreads having the normal number of 40 chromosomes. This was good enough to warrant continuing with these clones as they should be capable of generating germline chimaeras.

Table 6.3 Presence of ΔI506/7 mutation and chromosome counts of HR5 'hit' clones.

PLATE No.	HIT CLONE	MUTATION PRESENT ?	CHROMOSOMES (% normal mitotic spreads)
1	1.10	YES	100
2	2.13	YES	90
3	3.11	YES	100
4	4.3	YES	ND
	4.8	NO	ND
5	0	—	—
6	6.22	YES	ND
	6.24	YES	100
	6.25	YES	ND
	6.31	NO	ND
	6.38	YES	100
7	7.1	YES	ND
	7.7	NO	ND
	7.14	YES	100
	7.15	YES	90
8	8.5	YES	ND
TOTAL	**15**	**12 (90%)**	—

ND: Not done.

6.3.3 Conclusions.

Successful targeting of *Cftr* was again achieved, this time using the CGR8 cell line. A slightly higher transfection efficiency of 2.8 X 10^{-3}, and a much higher targeting frequency of 13.6% was obtained with this experiment than with HR3. This generated 15 'hit' clones, all but one of which appear to be the result of integration of a single vector molecule into the target site through the predicted homologous recombination mechanism. Multiple 'hit' clones were obtained from the same plates in some cases, suggesting the possibility that they arose from the same targeting event. However, due to the large numbers of clones analysed from some plates it is feasible that multiple but separate targeting events could have arisen on the same plate. The fact that in each case, one of the 'hit' clones arising on such a plate had lost the mutation, whilst other 'hit' clones had retained it, suggests that at least two independent events must have occurred on each plate. The ΔI506/7 mutation had been lost from 20% of these 'hit' clones leaving 12 'hit' clones suitable for taking forward to the 'run' stage. Chromosome analysis was carried out on some of these 'hit' clones, and revealed a normal or near normal chromosome number in all of those analysed.

6.3.4 Generating Transgenic Mice From Targeted 'Hit' Clones.

Targeted 'hit' clones from the HR5 targeting experiment were used in an attempt to generate chimaeras by techniques described in chapter 4, which could then be bred to generate mice which were homozygous for the disruption of *Cftr*. The production of these mice would be of interest for the following reasons:

(I) Before embarking upon the lengthy procedure of the 'run' stage of this strategy, it would be advantageous to ascertain whether these 'hit' clones were capable of contributing to the germ line of chimaeras and therefore transmitting a mutation. This was important, as the ultimate aim of this project was to produce mice from ES cell clones resulting from a 'run' event which have retained the mutation. Therefore it is important to ensure that the 'hit' clones have retained this ability to contribute to the germ line and generate chimaeric mice before they are taken through into the 'run'

stage. A problem with germ line transmission may arise through the presence of the *hsvtk* gene in the targeting vector of 'hit' clones, as it has been reported that *hsvtk* gene expression can disrupt sperm development causing male transgenic mice to be infertile (Braun *et al.* 1990). However, Ramirez-Solis *et al.* (1993) achieved germ line transmission of 'hit' clones which contained *hsvtk* when targeting the *Hoxb-4* locus, although this occurred at a lower frequency than with clones which did not contain the *hsvtk* gene.

(II) In addition to checking the germline competency of the 'hit' clones, the generation of mice with this targeted disruption and duplication of exon 10 would be of interest in terms of phenotype displayed by mice homozygous for this mutation. A mouse model has been generated for CF which has a targeted disruption and duplication of a region around exon 3, giving a structure similar to that of a targeted 'hit' event (O'Neal *et al.* 1993). The phenotype of this mouse is severe, as no wild type mRNA is generated, and the duplication of exon 3 causes a frameshift mutation as the exon is 109 base pairs in length. Another CF mouse model created using insertional mutagenesis (Dorin *et al.* 1992), also creates a duplication, this time part of exon 10. The phenotype of this mouse model is much milder than for the exon 3 duplication, due to the generation of a low level of wildtype mRNA by alternative splicing. This aberrant splicing pattern is caused by skipping of the mutant exon 10 by the splicing mechanism, and is thought to be due to a bias against the large exon formed by the inframe fusion of the partial exon 10 sequences to the plasmid sequences. It is difficult to predict the phenotype which might be shown by mice homozygous for the disruption and duplication of exon 10 generated by integration of the 'hit and run' targeting vectors. In this case, the duplicated exons will not cause a frame shift mutation and therefore mRNA and mutant protein should be generated. There may be a selection against mRNA with a duplicated exon which might not then be translated, resulting in a severe phenotype. It is also possible that due to the relatively normal nature of the duplicated exons, these will splice into the mRNA and be translated into protein. The presence of the ΔI506/7 might still be recognised by the 'quality control' mechanisms of the endoplasmic reticulum, leading to its failure to

process properly, thereby becoming mislocalised in the cell as in the human CF condition. Therefore, it would be interesting to see which of these possible scenarios would be the consequence of a mouse which is homozygous for the 'hit' mutation.

The targeted 'hit' clones 1.10 and 7.15 were used in an attempt to generate transgenic mice by the ES cell and morulae aggregation techniques described in chapter 4. These techniques had not previously been used in this lab, and the setting up and establishing of the required conditions have taken time. Although litters have been born, they have all originated from the morulae alone, and no chimaeras have been generated by this technique to date. This is probably a reflection of our lack of expertise in this area rather than a failure of the ES cells to generate mice. A different version of the aggregation chimaera technique is used successfully by Dr Andras Nagy. This uses a darning needle of a specific type (Hungarian) to create small depressions in a tissue culture dish, which forces the ES cells and the morulae into close contact without the need for phytohaemagluttinin (Nagy and Rossant, 1993). Our failure to generate chimaeras could be attributed to the excessive handling of the ES cells and the exposure to Phytohaemagglutinin entailed by our protocol, which might have a deleterious effect upon their pluripotency. These experiments are now being continued by another member of the group, primarily using the ΔF508 'hit' clones, but as yet, no mice have been generated.

6.4 DISCUSSION

This chapter has described the successful targeting of the murine *Cftr* gene with 'hit and run' targeting vectors as the first step towards the introduction of subtle mutations by the 'hit and run' targeting strategy. Clones have been generated which have a duplication of a region of exon 10 due to integration of a 'hit and run' targeting vector through homologous recombination. A large number of these targeted clones have retained the mutation carried by the vector and are suitable for taking through to the next 'run' stage where a negative selection against the vector will be imposed to enrich for cells which have excised the vector through intrachromosomal recombination.

6.4.1 Cell Lines

Two different ES cell lines were used for these targeting experiments, each derived from the same 129 strain of mouse. Both cell lines gave a relatively similar transfection efficiency of between 1-2 X 10^{-3}, but differed in the frequency at which *Cftr* was targeted. A targeting frequency of 4.6% was achieved with the E14 cell line, whereas a higher frequency of 13.6% was obtained using CGR8 ES cells. This is in contrast to the targeting frequencies obtained for different cultures of the same two cell lines, when targeting *Cftr* with a similar 'hit and run' vector (this time carrying the ΔF508 mutation) in our lab under similar conditions (Dickinson *et al.* 1993). In these experiments the more consistent targeting frequency of 8.5% and 10.9% were obtained for E14 and CGR8 cell lines respectively. The difference in targeting efficiency for the targeting experiments described in this chapter cannot be attributed to differences in transfection efficiency, as these were similar for both cell lines. The main difference between these experiments and those with the ΔF508 vector, is in the positive selection that was used. The experiments detailed here utilised Hygromycin as the positive selection, whereas the vector carrying the ΔF508 mutation used G418. The effectiveness of the Hygromycin positive selection is unlikely to vary in different cell lines, and any differences would be reflected in the transfection efficiencies. Differences in targeting efficiencies between identical experiments have also been reported by Ramirez-Solis *et al.* (1993), when targeting the *Hoxb*-4 locus, and by Koller *et al.* (1991) when targeting exon 10 of *Cftr* and another locus (*hprt*) in parallel. Koller *et al.* reported some correlation between increasing culture passage number and increased targeting efficiency. In this instance, it was the E14 cell line used here which had a higher passage number rather than the CGR8 cell line which gave the higher targeting frequency. However, the passage number of a cell line is only an empirical measure of its expansion and cannot be directly compared between cell lines. It is possible that cells are more receptive to targeting at different times, and the disparity in targeting efficiencies is more a reflection of this than of any true differences between the cell lines. It should be considered that the 'hit' clones obtained from targeting the E14 cell line had an abnormal karyotype which may be

representative of the culture used for the transfection. Therefore, this abnormal karyotype could be responsible for the depressed targeting efficiency in some way, especially as the E14 'hit' clones obtained at a much higher targeting frequency using the ΔF508 targeting vector did not have such an aberrant karyotype.

6.4.2 Targeting Frequencies.

The data generated by these targeting experiments, when combined with that of similar targeting experiments carried out in our lab by Fiona Kilanowski and Julia Dorin, has allowed some important observations to be made regarding the efficiency of targeting at this locus with insertion vectors (table 6.4, and Dickinson *et al.*, 1993).

6.4.2.1 Insertion versus replacement targeting vectors.

The targeting of genomic regions around exon 10 of *Cftr* has also been accomplished in other labs, (Koller *et al.* (1991)., Snouwaert *et al.* (1992)., Ratcliff *et al.* (1992)., van Doorninck *et al.* (1993)., and Deng et *al.* (1994)). The type of vector used and the targeting frequency reported varies considerably (figure 6.5). A high proportion of these targeting experiments have utilised replacement (Ω) type targeting vectors (described in section 1.2.4.1), and achieved largely very low targeting frequencies ranging from 0.04% (Koller *et al.* 1991), to 1.4% (Deng *et al.* 1994). This contrasts with the significantly higher targeting frequencies of 3.8-13.6% achieved in this and other labs using insertional (O) type targeting vectors. Despite the greater lengths of genomic homology incorporated into the replacement vectors, which are known to increase targeting efficiency (discussed later), the replacement vectors still targeted *Cftr* much less efficiently. The reported targeting frequencies for replacement vectors at this locus also varied enormously (by three orders of magnitude), whilst insertion vectors from two different labs target to an efficiency of within 10% of each other.

Table 6.4. Targeting efficiency at the *Cftr* locus using insertional vectors.

VECTOR	CELL LINE	5' HOMOLOGY (kb)	3' HOMOLOGY (kb)	LINEARISATION SITE	POSITIVE SELECTION	TARGETING FREQUENCY (%)
pIV3.5H	E14	0.4	3.1	*Asp 718*	G418	2.2
pHRNTΔF508	E14	0.4	3.9	*Asp 718*	G418	3.8
pHRNTΔF508	E14	1.2	3.1	*Hpa I*	G418	8.5
pHRNTΔF508	CGR8	1.2	3.1	*Hpa I*	G418	10.9
pHTHRΔI	**E14**	**1.2**	**3.1**	***Hpa I***	**HYGROMYCIN**	**4.6**
PHTHRΔI	**CGR8**	**1.2**	**3.1**	***Hpa I***	**HYGROMYCIN**	**13.6**

The bold type indicates targeting experiments described in this thesis.

Table 6.5. A comparison of reported targeting frequencies and vectors used to target similar regions around exon 10 of *Cftr*.

VECTOR TYPE	ISOGENIC DNA ?	LENGTH OF OVERALL HOMOLOGY (kb)	5' HOMOLOGY (kb)	3' HOMOLOGY (kb)	TARGETING FREQUENCY (%)	REFERENCE
Replacement	YES	7.8	0.7	7.1	0.04-0.3	Koller *et al.* (1991), Snouwaert *et al.* (1992)
Replacement	YES	5.5	1.2	4.3	0.13	Ratcliff *et al.* (1992)
Replacement	YES	12	~4	~8	1.0[a]	Deng *et al.* (1994)
Insertion	YES	3.5	0.4	3.1	2.2	Dorin *et al.* (1992)
Insertion	YES	4.3	0.4	3.9	3.8	Dickinson *et al.* (1993)
Insertion	YES	4.3	1.2	3.1	5-14	Dickinson *et al.* (1993)[b]
Insertion	NO	NA	NA	NA	0	van Doorninck *et al.* (1993)
Insertion	YES	NA	NA	NA	9	van Doorninck *et al.* (1993)

[a] The reported targeting frequency was adjusted to take into account the enrichment for targeted clones obtained by positive-negative selection. As no figure was stated for this enrichment factor, the commonly reported figure of 10% was used for this adjustment.
[b] This data includes some of the targeting experiments described in this thesis.
NA. This information was not described in the publication.

Increased targeting frequencies for insertion vectors when compared to replacement vectors have also been reported by another group when targeting a different mammalian locus. Hasty *et al.* (1991) conducted a comprehensive comparison of the efficiency at which similar insertion and replacement vectors targeted the *hprt* locus in ES cells. They consistently obtained a nine fold increase in targeting frequency when insertion vectors were used. This effect is disputed by Deng and Capecchi (1992), who conducted a similar comparison at a different region of the *hprt* locus. They found no difference in targeting efficiency between the two vector types when using isogenic targeting sequences in the vectors. However, when the genomic homologous sequences of the vectors were from a non-isogenic source, the targeting efficiency of the replacement vector was dramatically reduced. They suggest that replacement vectors are more sensitive to the presence of heterological sequences due the different mechanism of integration (described in detail in section 1.2.4.1). In order for a replacement vector to integrate into the genome through homologous recombination, a crossover event must occur on both arms of the vector (Hasty *et al.* 1991., Hasty *et al.* 1992., Deng and Capecchi, 1992., Hastings *et al.* 1993). If heterologous sequences are present on one arm of the vector, the efficiency of the crossover on that arm will be reduced, thereby reducing the efficiency of recombination for the entire vector. In contrast, insertion vectors only require a single crossover to occur in any region of the homology for efficient insertion of the whole vector into the endogenous sequences, and consequently will be less sensitive to the presence of any heterologies in the targeting sequences. Therefore, it could be argued that the decreased targeting efficiencies of replacement vectors observed by Hasty *et al.* are a consequence of their use of nonisogenic DNA in the targeting vectors. Thomas and Capecchi (1987), reported no significant difference when also targeting the *hprt* locus with both vector types. However, the insertion vector used in these targeting experiments had an unusual feature, as the Neomycin selection gene was present in the 3' arm of the targeting vector and so interrupted the length of homology. The amount of continuous homologous sequences is known to affect the targeting efficiency (discussed later), and could be responsible for suppressing the targeting efficiency of the insertional vector. The targeting frequencies of both vector

types have also been compared in the yeast *Saccharomyces cerevisiae* (Hastings *et al.* 1993). A two to three fold increase in targeting frequency was observed with insertion vectors in these experiments, and therefore is consistent with the same effect observed when targeting the *Cftr* locus in murine ES cells.

The proposed higher targeting efficiency of insertional vectors is thought to be a result of the mechanism of homologous recombination. As mentioned above (and described in detail in section 1.2.4.1), integration of an insertion vector is thought to occur through a single crossover point, whereas a replacement vector requires a crossover event to occur on both arms of the vector. The formation and stabilisation of two such recombination events may be less efficient than that of a single event, which would be reflected in the targeting efficiencies of the two vectors.

The higher efficiency targeting we and others have achieved at the *Cftr* locus using insertion (O) type targeting vectors is in agreement with that observed by one group at another locus and also in targeting experiments in yeast, but differs from the observations of some other labs. The 'double strand break repair' model for integration of DNA through homologous recombination (Orr-Weaver *et al.* 1981, Szostak *et al.* 1983) however proposes a different mechanism of integration for the two vector types, which leads to the prediction that the more simple insertion events are likely to be more efficient and consequently occur at a higher frequency.

6.4.2.2 Effect of homology carried by targeting vectors on targeting efficiency.

The targeting experiments carried out in our lab (table 6.4) have provided some important insights into the effect of the homologous sequences carried by the vector upon the targeting frequency at the *Cftr* locus. As all of the vectors used represent modifications of the pIV3.5H targeting vector, meaningful comparisons can be made between the effects of these modifications upon targeting frequency.

A simple increase in the length of homology to the target site of vector pHRNTΔF508 from 3.5 kb to 4.3 kb, resulted in an increase in targeting efficiency

from 2.2% to 3.8%. Although this is clearly not significant, some effect of increased homology resulting in a higher targeting frequency is expected. Rubnitz and Subramani (1984)., Thomas and Capecchi (1987)., Hasty *et al.* (1991)., Deng and Capecchi (1992)., and Berinstein *et al.* (1992), have all reported this phenomenon which has also been observed when targeting genes of the yeast genome (Rothstein 1991). Targeting frequency appears to have a strong dependence upon the length of homology between vector and target sequences. An increase in length of homology from 1.3 to 6.8 kb has been found to result in greater than a two hundred fold increase in the targeting efficiency of both insertion and replacement vectors (Hasty *et al.* 1991). Deng and Capecchi (1992) reported that a length of more than 14 kb of homology is needed before this effect diminishes. Although targeting vectors of different origin cannot be directly compared, it is interesting to note that out of all the replacement vectors used to target *Cftr* (table 6.5), the replacement vector which gave the highest targeting frequency also incorporated the greatest length of homologous sequences.

The distribution of the homologous sequences might also affect the efficiency of homologous recombination, as we observed an increase in targeting efficiencies when the site of the double strand break was moved by utilising a different restriction enzyme. Linearisation of the vector pHRNTΔF508 at the *Asp 718* restriction site generated 0.4 kb of homology 5', and 3.9 kb of homology 3', which resulted in a targeting frequency of 3.8%. Utilisation of the *Hpa I* restriction site gave a more even distribution of homology with 1.2 kb 5', and 3.1 kb 3', and increased the targeting frequency significantly from 3.8% to 8.5%. This effect could be attributable to a more even distribution of homology either side of the double strand break, or the creation of more 'recombinogenic' free ends. Both Hasty *et al.* (1991) and Berinstein *et al.* (1992) found that targeting efficiencies were determined by the total length of homology of vector sequences to the target site, rather than distribution of genomic sequences. A complete failure to target *Cftr* was reported by van Doorninck *et al.* when a *Nhe I* site was used to linearise the insertion vector, however, the same vector gave a targeting frequency of 9% when cut with *Nsi I*. The effect of the different

restriction sites on the distribution of homology was not given, but van Doorninck *et al.* suggested that this is due to a property of the sequences around the site of linearisation. Therefore the increased frequency observed when the targeting vector was linearised with *Hpa* I is probably a property of the double strand break created by the restriction site.

6.4.2.3 Comparison of 'hit and run' vectors.

The 'hit and run' targeting vector pHRNTΔF508 differs from the vector used in the targeting experiments described in this thesis, only by the mutation carried and the selection genes used. Vector pHRNTΔF508 carries a gene for resistance to the positive selection G418, whereas pHTHRΔI carries the gene for Hygromycin resistance. Despite this, the targeting frequencies obtained with these vectors are remarkably similar, except for the targeting experiment HR3 in which the targeting frequency appears to have been depressed, perhaps due to the abnormal chromosome number. Therefore, the positive selection regime used does not appear to affect the targeting frequencies achieved.

6.4.2.4 Source of DNA.

One factor which is known to have a significant effect on targeting frequency is the source of genomic DNA. If the homologous sequences present on the vector are not isogenic with the ES cells, that is to say from the same mouse strain, the small sequence heterologies present have been found to have a profound negative effect upon targeting efficiencies (te Riele *et al.* 1992, van Deursen and Wieringa 1992, Deng and Capecchi 1992, Wurst *et al.* 1994). The magnitude of this effect varies from locus to locus. This was not an issue with our targeting vectors as isogenic DNA was used throughout, however, this factor does appear to affect targeting frequency at the *Cftr* locus as demonstrated by the failure of van Doorninck *et al.* (1993) and O'Neal *et al.* (1993) to target this locus when using a non-isogenic targeting construct.

6.4.3 Structure of Insertion Events.

All but one of the targeted 'hit' clones obtained in the targeting experiments described in this thesis were the result of a simple integration of a single vector into the target site in the predicted manner. This is in line with the common observation that integration of a vector into a target site through homologous recombination is rarely accompanied by insertion of another copy elsewhere in the genome. This phenomenon was described by Reid *et al.* (1991) when attempting to cotransfect two unrelated insertion vectors into ES cells as a method of enriching for a non-selectable homologous recombination event. They found that although non-targeted cotransformations occurred frequently, the homologous recombination of a vector with the target site was rarely (in only 4% of cases) accompanied by integration of a second DNA molecule. In addition, DNA that had integrated through homologous recombination was usually present as a single copy, whereas DNA which had integrated through non homologous means was frequently present as concatemers at multiple sites throughout the genome. The integration of a single copy of DNA in the predicted manner is a frequent observation of targeting with insertion vectors, and contrasts with the aberrant integration patterns reported for replacement vectors (Hasty *et al.* 1992., Thomas *et al.* 1992., Hasty *et al.* 1992., Davis *et al.* 1992., Zhang *et al.* 1994). Integration of replacement vectors often occurs through mechanisms such as insertion, and is thought to be due in some cases to the recircularisation of the vector, or the formation of vector concatemers prior to integration. The apparent low occurrence of such a phenomenon with insertion vectors is thought again to be a reflection of the different methods of integration for the two vector types. Integration of an insertion vector is thought to occur through invasion of one of the free ends of the double strand break into the homologous region of the target site, as predicted by the double strand gap repair hypothesis (Szostak *et al.* 1983), described in section 1.2.4. This displaces a strand of DNA which is then available for strand invasion by the other free end. Both ends of the double strand break become rapidly associated with the target site DNA and are not available for recircularisation or concatemer formation. Integration of replacement vectors however is thought to occur via a double crossover event, which is thought to be less efficient than for an insertion

event, leaving the free ends of the double strand break exposed for longer periods to the ligation activities of the cell.

6.4.4 Loss Of Mutations During Integration.

The loss of the ΔI506/7 mutation in 33% of the targeted HR3 'hit' clones and 20 % of HR5 'hit' clones is a phenomenon which has been widely reported (Valancius and Smithies 1991a., Valancius and Smithies 1991b., Hasty *et al.* 1991., Hasty *et al.* 1992., Deng and Capecchi 1992., O'Neal *et al.* 1993., Rubinstein *et al.* 1993., Ramirez-Solis *et al.* 1993., Wu *et al.* 1994). The frequency of loss of a mutation has been reported to be related to its proximity to the double strand break used to linearise targeting vectors (Valancius and Smithies 1991., Deng and Capecchi 1993), an observation that has also been made in yeast targeting experiments (Orr-Weaver *et al.* 1988., Rothstein 1991). Targeting experiments conducted in our lab using a very similar 'hit and run' targeting vector to introduce the ΔF508 mutation, experienced a similar frequency of mutation loss of 33%. As the position of the ΔF508 mutation from the double strand break differed by only 3 base pairs from the position of the ΔI506/7 mutation, the comparable frequency of mutation loss is therefore consistent with this proposed relationship of frequency of mutation loss to distance from the double strand break. When targeting the same region of *Cftr* with a vector in which the ΔF508 mutation was located only 620 base pairs away from the double strand break, Van Doorninck *et al.* (1993) found that 84% of 'hit' clones had lost the mutation. The observations at this locus are consistent with those observed by others at different loci. Wu *et al.* (1994) report a loss of a mutation located 670 base pairs from a double strand break in 59% of clones when targeting the Collagen type I gene in ES cells. The frequency of mutation loss was investigated by Deng and Capecchi (1993) at the *hprt* locus by placing the mutation at varying distances from the double strand break. When this distance was only 0.7 kb, the mutation was lost in 74% of targeted clones, whereas a distance of 4.2 kb resulted in mutation loss in only 5% of clones. The loss of the ΔI506/7 mutation in 20-33% of the 'hit' clones when situated 2 kb from the double strand break is consistent with the rate of mutation loss compared to proximity to the double strand break relationship observed at other loci.

This loss of heterologous sequences appears to be largely a feature of homologous recombination of insertion vectors, and has been reported to occur extremely rarely in cells targeted with replacement vectors (Deng and Capecchi 1993). It has been suggested that this is a reflection of the different integration pathways utilised by insertion and replacement vectors during homologous recombination (Hasty *et al.* 1992). Insertion vectors are thought to integrate via the double strand gap repair pathway (Orr-Weaver *et al.* 1981., Szostak *et al.* 1983., Hasty *et al.* 1992., Deng and Capecchi 1993). During the integration process, the free ends of the double strand break are thought to be degraded by the action of an exonuclease and the double strand gap enlarged before being repaired using the target site as the template (Valancius and Smithies, 1991, Hasty *et* al. 1992, Deng and Capecchi 1993, Kumar and Simons 1993) Targeting experiments incorporating double strand gaps rather than breaks have been reported to efficiently and precisely target a locus, with the missing homology from the double strand gap being 'filled in' during integration of the vector (Valancius and Smithies 1991). This phenomenon has been exploited by Smithies and Kim (1994) to create functional duplications of the Angiotensin gene, by targeting the gene with an insertion vector bearing homology to the most 5' and 3' regions. The 8 kb gap in the genes homology was repaired during the integration of the vector to create a complete tandem duplication of the gene. Heterologous sequences present at the site of the double strand break are invariably lost during homologous recombination, consistent with the occurrence of exolytic enlargement of the double strand break prior to integration (Hasty *et* al. 1992, Kumar and Simons 1993). The correlation of mutation loss with proximity to the double strand break is therefore likely to be a reflection of the degree of exonuclease gap enlargement required for removal of the mutation.

The observation of mutation loss and the proposed mechanism has consequences upon the design of a 'hit and run' targeting vector. To minimise mutation loss through exonuclease gap enlargement, the mutation should be positioned away from the site used to linearise the vector. However, to maximise the region through which

recombination can occur in the 'run' stage to excise the wild type copy, the mutation should be placed at a distance from the plasmid sequences. Therefore a compromise must be reached between the distance from the double strand break which should result in retention of the mutation at a reasonable frequency, and the homology available for recombination between the mutation and the plasmid sequences (Hasty and Bradley, 1993). In hindsight, although the configuration of the 'hit and run' targeting vector pHTHRΔI used in these targeting experiments resulted in a sufficiently high frequency of mutation retention, the length of homology available for intrachromosomal recombination during the 'run' stage favours loss of the mutation.

Some groups have reported that the position of a mutation in a targeted 'hit' event can occur in an orientation other than that predicted (Hasty *et* al. 1991, O'Neal *et al.* 1993, Wu *et al.* 1994). In a small proportion of clones, the mutation has been found to be present in the opposite duplicate to that intended, and is thought to be a consequence of migration of the Holliday junction across the mutation site to resolve on the other side during the recombination process. Alternatively, the mutation has also been found to be present in both duplicated sequences at an equally low frequency and is also thought to be due to migration of a Holliday junction, but in this case preceded by mismatch heteroduplex repair in favour of the mutant sequences (Hasty and Bradley 1993). The generation of such a clone with the mutation present in both duplicated regions would be advantageous for use in the 'run' stage, as excision of the vector should always result in the retention of the mutation in a 'run' clone. Unfortunately, none of the 'hit' clones generated in these targeting experiments had two copies of the mutant allele (as determined by Southern blot analysis, (figures 6.5 and 6.7) and so did not have this advantage.

The observations made during the targeting of the *Cftr* locus are therefore in general agreement with those made by others when targeting the same and different loci. The data generated supports the theory that insertion vectors target more reliably and at a higher frequency than replacement vectors. The targeted 'hit' clones generated by

these targeting experiments are suitable for use in the 'run' stage of the 'hit and run' targeting strategy.

CHAPTER 7

SELECTION FOR THE 'RUN' EVENT

7.1 INTRODUCTION

The previous chapter described the generation of 'hit' clones in which the 'hit and run' targeting vector pHTHRΔI506/7 had integrated into the target site of *Cftr* through homologous recombination. This created a duplication of the target site with, in the majority of 'hit' clones, the ΔI506/7 mutation present in one of the duplicates (figure 7.1). Twelve such 'hit' clones possessed the ΔI506/7 mutation, and were suitable for use in the 'run' stage.

The second step of the 'hit and run' targeting strategy, the 'run' step, entailed the imposition of a negative selection onto targeted 'hit' clones to select against the integrated vector. It was envisaged that a certain proportion of cells would excise the integrated vector through intrachromosomal recombination within the duplicated sequences, and survive the negative selection regime due to the concurrent loss of the Hygromycin-*tk* fusion gene along with the vector sequences. Depending upon the actual position of the crossover, the excision recombination event could have one of two consequences: (a) recombination anywhere within the endogenous 3' duplicated region would result in loss of the integrated vector and a reversion back to a wildtype (non-targeted) status. (b) recombination within the 5' duplicated sequences would excise the endogenous wildtype exon 10 along with the vector, leaving the mutant exon in its place. The ΔI506/7 mutation in the mutant exon would then be present as the only alteration to the target site.

7.2 RUN2 'RUN' EXPERIMENT

7.2.1 Method

The 'hit' clone 1.10 was expanded on a PEF feeder layer and 9.2 X 10^7 cells used for selecting 'run' events. These cells were seeded at different densities, and in the presence or absence of feeder layers, to determine the influence of these different conditions upon the effectiveness of the selection and perhaps frequency of the 'run' event. These conditions are detailed in table 7.1.

Table 7.1 The conditions used for seeding cells for RUN2.

TOTAL No. CELLS (x 10^6)	SEEDING DENSITY (No. cell/plate x 10^6)	TOTAL No. PLATES	PEF FEEDER LAYER ?
12	2	6	YES
40	2	20	NO
40	4	10	NO
92	—	36	—

Twenty four hours after the cells were plated out, Gancyclovir negative selection was imposed. The minimum concentration of Gancyclovir required to kill cells expressing the Hygromycin-*tk* fusion gene had been previously established on a targeted 'hit' clone by the method described in section 2.1.9. In order to determine whether this concentration of 2.5 μM, or a slightly higher concentration would be more effective, the selection was imposed on half of the plates at a concentration of 2.5 μM, and the other half at 3.5 μM. The Gancyclovir-containing medium was replaced every two days. Once visible, clones were picked and grown up to generate cells for frozen stocks and DNA analysis.

7.2.2 Results

7.2.2.1 Clone numbers.

After six days of Gancyclovir selection, clones were visible and ready for picking from those plates in which the cells had been plated at a density of 2 X 10^6 cells per plate in the absence of a PEF feeder layer. Cells which had been plated at the higher density of 4 X 10^6, also in the absence of PEFS, became confluent before the selection took effect. A much longer period was required before these cells began to die in Gancyclovir, and when this did occur, a very small number of Gancyclovir resistant clones were obtained from only one of the ten plates. The plates in which the cells had been plated onto a PEF feeder layer also became confluent before the selection was effective, but these plates generated a very large number of Gancyclovir resistant clones.

The total number of clones present on each plate, and the number of clones picked from each are detailed in table 7.2, and summarised in table 7.3. A total of 330 Gancyclovir resistant clones were picked from the clones arising in this experiment. Although a much higher number of clones arose on those plates in which cells had been selected upon a PEF feeder layer, fewer clones were picked from these plates as the high clone density meant it was difficult to distinguish and pick individual clones. Very few clones arose on the plates where cells had been plated at the higher density of 4×10^6 cells per plate, whereas plates seeded at the lower density of 2×10^6 cells per plate under identical conditions generated a substantial number of Gancyclovir resistant clones. A difference was observed between the frequency of Gancyclovir resistant clones obtained with the two different concentrations used, with the higher concentration generating a smaller number of clones.

Table 7.2 The numbers of clones arising in experiment RUN2.

	SEEDING DENSITY (No. cells/plate x 10^6)	PEF FEEDER LAYER ?	CONC. GANC (μM)	TOTAL No. CLONES	No. CLONES PICKED
	2	NO	2.5	33	20
	2	NO	2.5	28	28
	2	NO	2.5	26	16
	2	NO	2.5	52	32
	2	NO	2.5	21	16
	2	NO	2.5	28	6
	2	NO	2.5	33	11
	2	NO	2.5	19	8
	2	NO	2.5	NC	0
	2	NO	2.5	NC	0
TOTAL	—	—	—	**240**	**137**
	2	NO	3.5	35	26
	2	NO	3.5	29	15
	2	NO	3.5	35	26
	2	NO	3.5	17	10
	2	NO	3.5	19	12
	2	NO	3.5	25	8
	2	NO	3.5	NC	0
	2	NO	3.5	NC	0
	2	NO	3.5	NC	0
	2	NO	3.5	NC	0
TOTAL	—	—	—	**160**	**97**
	4	NO	2.5	0	0
	4	NO	2.5	0	0
	4	NO	2.5	0	0
	4	NO	2.5	0	0
	4	NO	2.5	0	0
TOTAL	—	—	—	**0**	**0**
	4	NO	3.5	12	8
	4	NO	3.5	0	0
	4	NO	3.5	0	0
	4	NO	3.5	0	0
	4	NO	3.5	0	0
TOTAL	—	—	—	**12**	**8**
	2	YES	2.5	~1000	24
	2	YES	2.5	~1000	23
	2	YES	2.5	~1000	0
TOTAL	—	—	—	**~3000**	**47**
	2	YES	3.5	~500	24
	2	YES	3.5	~500	17
	2	YES	3.5	~500	0
TOTAL	—	—	—	**~1500**	**41**

NC. Not counted

Table 7.3 A summary of the number of Gancyclovir resistant clones arising in experiment RUN2.

SEEDING DENSITY (No. cells/plate x 10^6)	PEF FEEDER LAYER ?	CONC. GANC (μM)	FREQUENCY $GANC^R$ CLONES (x 10^{-5})	No. CLONES PICKED
2	NO	2.5	1.5	137
2	NO	3.5	1.3	97
4	NO	2.5	0	0
4	NO	3.5	0.06	8
2	YES	2.5	50	47
2	YES	3.5	25	41

7.2.2.2 Screening HR5 RUN2 clones for loss of the 'hit and run' targeting vector.

Unfortunately a large number of the clones which had been picked became contaminated before DNA and frozen stocks were generated, and so were unavailable for analysis. From the 330 clones picked, only 89 made it through to the analysis stage.

The DNA from the 'run' clones was analysed by Southern blot. Initially, the same combination of *Xba I-Sal I* double digests probed with the 0.6EH probe used to identify 'hit' clones in which the vector had integrated into the target site (section 6.2.2.2), was used to screen 'run' clones for those in which the vector had excised. However, the pattern of bands obtained with these digests was confusing, as the *Sal I* restriction enzyme appeared to have not cut the DNA to completion. Despite taking measures to improve the degree of cutting with this enzyme, such as using a new batch of enzyme, and cleaning up the DNA by Ammonium acetate-Ethanol precipitation, the partially digested bands were still obtained. However, the *Xba I* restriction enzyme had cut to completion, indicating that the quality of the DNA could not be too low.

In order to determine if the vector was still present in these clones, another diagnostic Southern blot was conducted, this time on DNA digested with the restriction enzyme *Bgl I*. These digests were blotted and probed with the same 0.6EH probe used for the *Xba I-Sal I* digests. In this instance, the probe will hybridise to two bands in wildtype clones, and three in 'hit' clones, due to the presence of a *Bgl I* site in the genomic fragment used as the probe (figure 7.1). In wildtype clones a 23 kb fragment representing the *Bgl I* fragment 5' to the site in the probe, and a 9 kb fragment representing the 3' *Bgl I* fragment, will hybridise to this probe. In 'hit' clones the 23 kb fragment detected by this probe is reduced to 3.8 kb by cutting at the additional *Bgl I* sites present in the vector. Therefore, 'hit' clones are identified by the presence of a 3.8 kb mutant band in addition to the two wildtype bands, and this mutant band would be absent in 'run' clones which had excised the vector. Furthermore, in 'hit' clones, the 9 kb band should be twice the intensity of the other wildtype band and the 3.8 kb mutant band as there should be twice the number of copies present. The 23 kb band and the 3.8 kb mutant band should be of equal dosage as there should be only one copy of each present. Due to its large size however, the 23 kb wild type band often does not transfer as efficiently as the smaller bands and so this dosage effect may not always be evident.

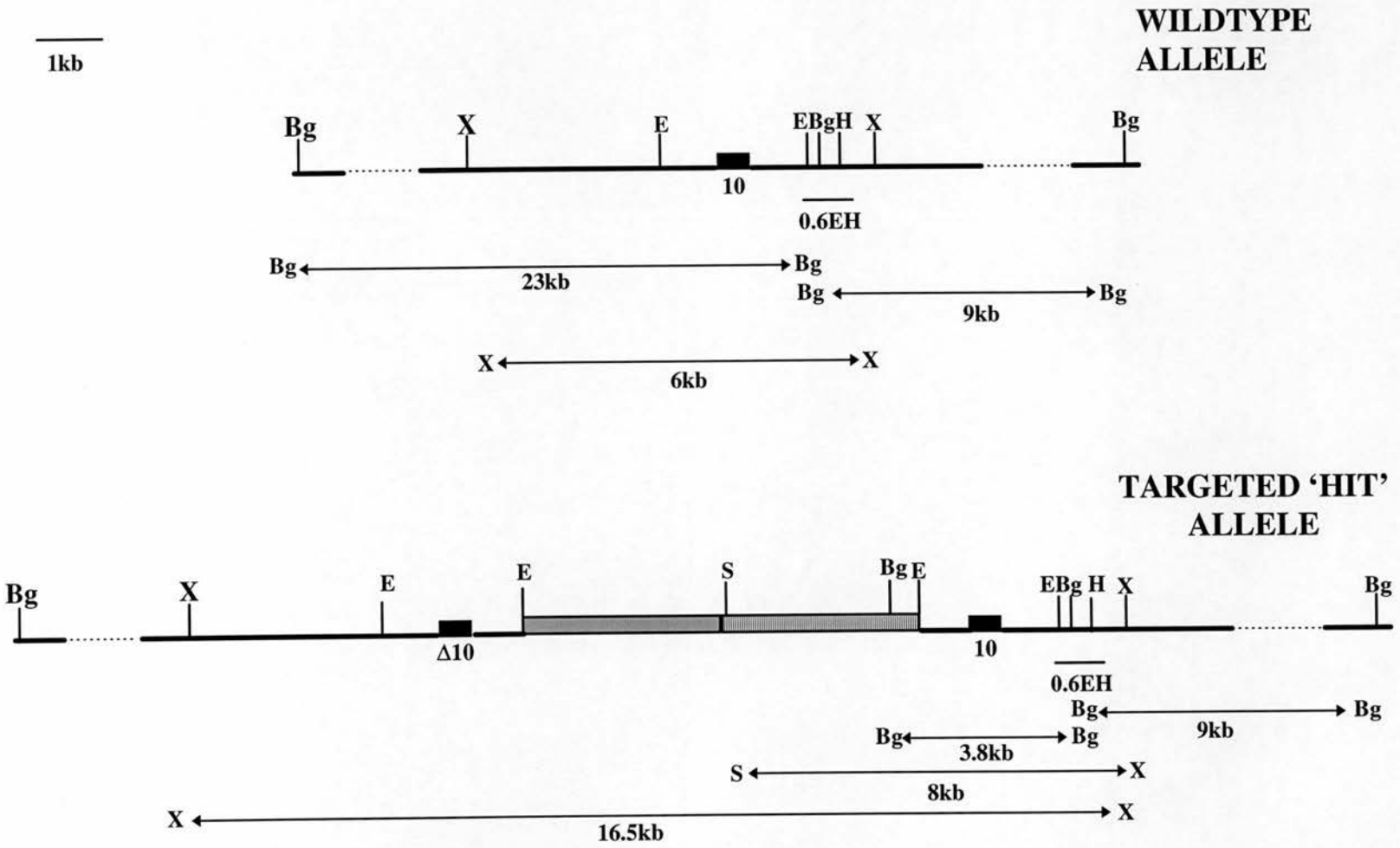


Figure 7.1 The structure and restriction map of targeted and wildtype alleles.
The solid black line represents *Cftr* genomic sequences, the black rectangle exon 10, and the position of the ΔI506/7 mutation is marked by Δ. Plasmid sequences are shown as a cross hatched region, and the Hygromycin-*tk* fusion gene by vertical hatching. Restriction sites are *Bgl I* (Bg), *Eco RI* (E), *Hind III* (H), *Sal I* (S), and *Xba I* (X). The site of hybridisation of the 0.6EH probe is indicated on both the wildtype and mutant allele.

When analysed by this digest and probe combination, 100% of the RUN2 clones were identical to the HR5 1.10 'hit' clone which had not undergone Gancyclovir selection, as they displayed the 3.8 kb mutant band as well as the wildtype bands (figure 7.2). There was no obvious size difference between the mutant band of the non-selected 'hit' clone and that of the 'run' clones, suggesting that there were no large deletions or rearrangements which could account for inactivation of *tk* gene.

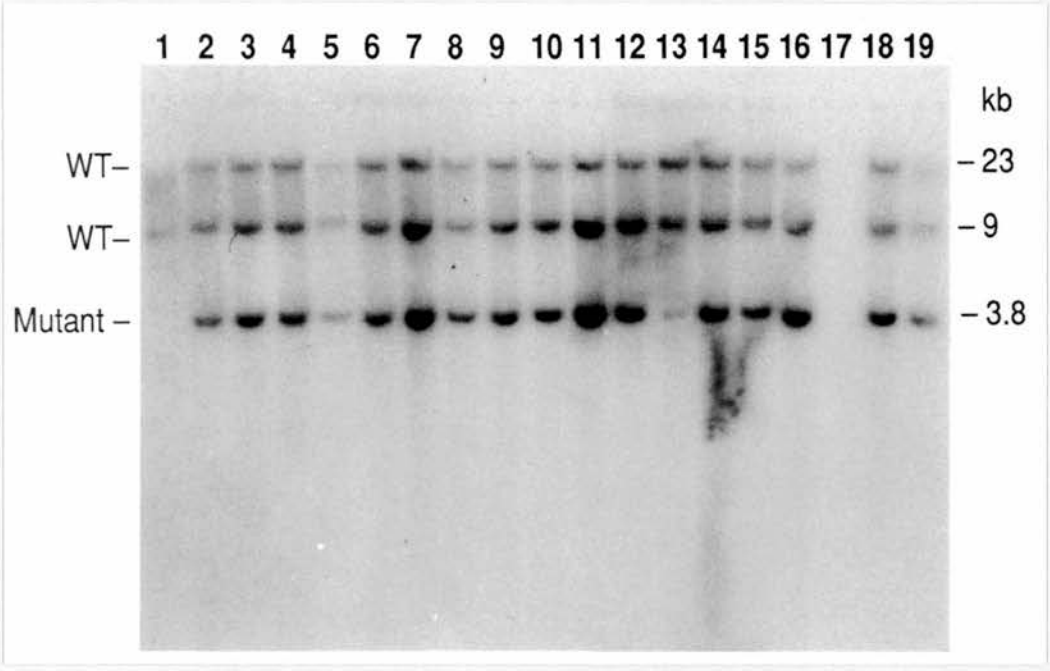


Figure 7.2 Southern blot analysis of RUN2 Gancyclovir resistant clones.
A Southern blot of *Bgl I* digests was probed with the 0.6EH genomic probe. Lane 1 shows the two wildtype bands of 23 kb and 9 kb, obtained with DNA from non-transfected CGR8 ES cells. (The bands are less distinct in this lane due to reduced DNA loading). Lane 2 represents DNA from the 'hit' clone 1.10 which had not been exposed to Gancyclovir selection, and displays the 3.8 kb band of the mutant targeted allele in addition to the two wildtype bands. Lanes 3 to 19 show bands which were typical of all RUN2 clones, and identical to those of the non-selected 'hit' clone, except for lane 17 which does not contain any DNA.

7.2.2.3 Screening RUN2 clones for Hygromycin resistance.

All of the 'run' clones had appeared to survive Gancyclovir selection despite containing the integrated vector. In order to determine if the Hygromycin-*tk* fusion gene was still functional, the positive Hygromycin selection used to obtain the 'hit' clones, was imposed upon some of the 'run' clones.

Twenty six 'run' clones were grown in medium containing Hygromycin at the same concentration used to select the 'hit' event. Immediately after imposition of Hygromycin selection, the cultures of 'run' clones appeared to die back, as was the case with non-transfected, wild type CGR8 cells used as a control. However, in contrast to wild type cells, after two weeks of selection the cultures from the 'run' clones had grown back to give healthy looking, fully Hygromycin resistant cultures. The 'hit' clone 1.10, from which these 'run' clones were derived, did not exhibit any cell death in Hygromycin selection and appeared to be completely Hygromycin resistant.

7.2.3 Conclusions

This experiment was not successful in generating true 'run' clones in which the vector had excised, as all of the Gancyclovir resistant clones obtained still appeared to contain the intact 'hit and run' targeting vector. Useful information however, was obtained regarding the effectiveness of Gancyclovir selection under different culture conditions. Despite the failure to obtain clones in which the vector had excised, this selection was apparently effective, as cell death occurred on all plates, and in most cases generated Gancyclovir resistant clones. The different conditions in which the cells were seeded appeared to affect the number of Gancyclovir resistant clones obtained.

Growing the cells on PEF feeder layers generated far more resistant clones than on the equivalent non-feeder layer plates. The mechanism of Gancyclovir resistance however, did not appear to differ between clones arising in the presence or absence of a PEF feeder layer, and therefore the higher number of clones generated was probably

a reflection of the higher plating efficiency displayed by CGR8 ES cells when grown on a PEF feeder layer.

The density at which the cells were seeded also affected the frequency at which resistant clones were obtained, with the higher seeding density of 4 X 10^6 cells per plate appearing to give a lower frequency of Gancyclovir resistant clones. This negative effect of higher seeding densities is probably due to the 'bystander' or 'cross-feeding' effect (Hooper, 1987.; Hasty and Bradley, 1993). This is a problem known to occur with negative selection regimes like the *tk*-Gancyclovir or *hprt* system, in which a non-toxic selection agent is metabolised to a cytotoxic analogue by cells expressing the negative selection gene. The propensity of ES cells to aggregate, and high seeding densities, facilitate cross feeding between cells. Under these conditions, a cell expressing *tk* which will die in Gancyclovir selection, can pass the cytotoxic base analogue to a non-expressing cell, causing it to also die in the selection even though it is not expressing the *tk* gene. This phenomenon has even been exploited in gene therapy studies where administration of Gancyclovir to rats with experimentally induced cerebral gliomas resulted in complete regression of tumours which had as few as 20% of cells transfected with the *hsvtk* gene (Culver *et al.* 1992). The potential for this effect to occur during selection for a 'run' event is great, as the majority of cells will be expressing *tk* and a non-expressing cell is likely to be surrounded by a large number of potentially lethal expressing cells. This effect can be overcome by seeding cells at low densities so that the contact between *tk* expressing and non-expressing cells is minimised. From our results it appears that a seeding density of 2 X 10^6 cells per 100 mm plate is sufficiently low to allow the survival of non-expressing cells, whereas the density of 4 X 10^6 is too high. The failure of the majority of plates seeded at this higher density to generate Gancyclovir resistant clones is therefore likely to be a result of the 'bystander effect'. In light of this, seeding densities for all future 'run' experiments were no higher than 2 X 10^6 cells per plate.

The concentration at which the Gancyclovir selection was imposed also seems to affect the number of resistant clones obtained, with more arising from selection at the lower concentration of 2.5 μM. It is difficult to ascertain the effectiveness of either concentration, as none of the Gancyclovir resistant clones were 'true runs' (i.e. they still contained the *tk* gene), and the mechanism by which they survived the selection is unknown. It is possible that the *tk* gene had been inactivated by some mechanism and that the Gancyclovir selection was selecting clones which did not express the *tk* gene. There does not appear to be any differences between the 'false positive' Gancyclovir resistant clones obtained at each concentration, and therefore both concentrations appear to have selected for the same event. If this is the case, and the clones obtained do not express *tk*, the lower numbers of resistant clones obtained at the higher concentration of Gancyclovir could be a reflection of a non-specific toxicity of Gancyclovir which might be seen at high concentrations. Based upon this information, the lower concentration of 2.5 μM Gancyclovir was used in all future experiments.

The demonstration by Southern blot analysis that the 'hit and run' targeting vector was still present in the 'run' clones showed that they had survived the negative selection by an unexpected mechanism. The absence of any detectable difference in the structure of the targeted allele of these 'run' clones when compared to that of the 'hit' suggests that they still contain the intact, integrated 'hit and run' targeting vector. It was envisaged that recombination within the duplicated sequences would result in excision of the vector and loss of the *tk* gene, thereby rendering the cells resistant to Gancyclovir selection. Instead, some other mechanism resulting in loss of *tk* gene expression appeared to be at work which did not alter the gene structure discernibly. Similar experiments targeting different loci have reported a range of frequencies for the vector excision event. Hasty *et al.* (1991) reported a frequency of 3.8×10^{-3} cells at the *hprt* locus, and a lower frequency of 4.3×10^{-6} cells at the *Hox 2.6* locus. An even lower frequency of 8×10^{-7} cells was observed by Valancius and Smithies (1991) at the *hprt* locus, implying that the frequency of an excision event is highly variable both between and within genes, and could be specific to the recombinogenic

nature of a particular region. Therefore it is difficult to predict the frequency at which this excision event might occur within the duplicated regions of the 'hit' clones in these experiments.

An investigation into the Hygromycin resistance of these clones also gave some unexpected results. As both Hygromycin resistance and Gancyclovir sensitivity are encoded by the same fusion gene, any mechanism which inactivates the expression of the *tk* gene would be expected to also inactivate expression of the whole fusion gene, thereby also rendering the clones Hygromycin sensitive. The Gancyclovir resistance of the 'run' clones which contain the intact 'hit and run' targeting vector suggests that the *tk* gene must have been inactivated in some way. When this was tested by putting the 'run' clones back onto Hygromycin selection, the clones initially appeared to be Hygromycin sensitive and began to die, but then grew back again. This suggests that either Hygromycin selection was also inactivated in the clones, but they eventually were able to overcome this inactivation by some means, or that a small proportion of the cells were Hygromycin resistant, and it was these that increased in numbers over time. In either case, the mechanism by which these 'run' clones were able to survive Gancyclovir selection was not known.

The results from this experiment are in contrast to those obtained in similar experiments conducted in parallel in our lab. Selection for 'run' events on targeted 'hit' clones containing the ΔF508 mutation which were comparable to those 'hits' used in these experiments, found that the majority of Gancyclovir resistant clones obtained after selection appear to be true 'run' events in which the vector has excised. In all cases, excision of the vector resulted in reversion back to wildtype status. A few 'false positives' were seen in these experiments where clones surviving the selection had not excised the vector, but in contrast to this experiment, these were by far the minority of Gancyclovir resistant clones obtained. The discrepancy between the two experiments was difficult to reconcile considering the similarity of the two 'hit and run' targeting vectors.

7.3 RUN3 'RUN' EXPERIMENT

The previous 'run' experiment was unsuccessful in generating any clones which had excised the 'hit and run' targeting vector. This experiment was repeated again with the same 'hit' clone, but was conducted in a slightly different manner.

7.3.1 Method

Cells from a passage 26 culture of the 'hit' clone 1.10 were used to select for a 'run' event. These cells were seeded at two different densities, 1 X 10^5, and 1 X 10^6 cells per 100 mm tissue culture plate. Lower seeding densities were used than in the previous experiment, in an attempt to minimise the loss of non-expressing cells through the 'cross-feeding' effect. All cells for selection of the 'run' event were grown upon PEF feeder layers during selection as this was found to increase the number of Gancyclovir resistant clones obtained in the previous experiment.

In order to obtain some estimate of the rate of vector excision, the Gancyclovir selection was added to a set of plates seeded at both densities, at daily intervals after seeding. Therefore, on the day of plating, Gancyclovir was added to a set of plates containing cells at both seeding densities which were labelled 'day 0'. Twenty four hours after plating out, selection was added to an identical set of plates labelled 'day 1', and so on until seven days after plating, selection was added to the last set of plates labelled 'day 7'. This regime was to be repeated with cells which had been grown on Hygromycin selection prior to Gancyclovir selection, and results from the two experiments compared.

Control plates were also included to determine that the Gancyclovir selection was working efficiently and killing those cells which were expressing the Hygromycin-*tk* fusion gene. For this, cells were plated at one of the densities used for selecting the 'run' event (1 X 10^5 cells per 100 mm plate), and maintained in Hygromycin selection to ensure that the fusion gene was being expressed, as demonstrated by their resistance to Hygromycin. Gancyclovir selection was then added to one plate per day sequentially, as with those plates for selection of the 'run' event. Clones expressing

the fusion gene should not be able to survive both Hygromycin selection and Gancyclovir selection when imposed simultaneously, and consequently the Hygromycin resistant cells should die on addition of Gancyclovir. Any cells able to specifically inactivate the *tk* gene expression of the fusion gene (e.g. by a point mutation in *tk)* should be detected on the control plates. As PEFS are not resistant to Hygromycin, these cells were selected in the absence of PEF feeder layers, unlike selection for the 'run' event.

Gancyclovir was used at the concentration of 2.5 μM. The medium and appropriate selection was replaced every three days, and once visible, the clones were picked and grown for frozen stocks and DNA analysis.

7.3.2 Results

7.3.2.1 Clone numbers.

Clones were visible on average eight days after addition of Gancyclovir selection. The plates in which Gancyclovir was added at the later days after plating were confluent by the time the selection was imposed. The numbers of clones obtained on each plate are detailed in table 7.4. No clones were visible on the control plates, indicating that the selection was working efficiently, and that cells expressing the fusion gene did not survive the Gancyclovir selection. A greater number of clones were obtained at the higher seeding density under all conditions in this experiment, unlike experiment RUN2 (section 7.2). The number of Gancyclovir resistant clones did increase with delayed Gancyclovir addition, although after day 2 it was difficult to distinguish individual clones and so it was not possible to determine clone numbers and establish if this effect continued throughout the time course. For this reason the number of clones arising on each plate are only detailed up to day 3 in table 7.4.

Table 7.4 The number of Gancyclovir resistant clones obtained in experiment RUN3.

PLATE No.	TIME OF GANC ADDITION (Days after plating)	SEEDING DENSITY (No. cells/plate x 10^6)	TOTAL NUMBER OF CLONES
1	0	0.1	80
2	0	0.1	120
TOTAL	—	—	**200**
3	0	1.0	400
4	0	1.0	360
TOTAL	—	—	**760**
5	1	0.1	224
6	1	0.1	184
TOTAL	—	—	**408**
7	1	1.0	1000
8	1	1.0	1400
TOTAL	—	—	**2400**
9	2	0.1	544
10	2	0.1	352
TOTAL	—	—	**896**
11	2	1.0	TMTC
12	2	1.0	TMTC
TOTAL	—	—	**TMTC**
13	3	0.1	TMTC
14	3	0.1	TMTC
TOTAL	—	—	**TMTC**
15	3	1.0	TMTC
16	3	1.0	TMTC
TOTAL	—	—	**TMTC**

TMTC. Too many clones to count. Individual clones not distinguishable.

7.3.2.2 Screening HR5 RUN3 clones for vector excision.

A total of 112 RUN3 clones were picked and grown up for DNA analysis. Microbial contamination was a problem again however, and only 8 clones were available for analysis. These eight clones were analysed by the same method used to identify targeted 'hit' clones, in which a Southern blot of *Xba I-Sal I* double digests was probed with the 0.6EH genomic probe (6.2.2.2). This detects a wildtype band of 6

kb, and a band of 8 kb which represents the integrated vector. A true 'run' event, in which the vector had excised, would be identified by the absence of the 8 kb mutant band with only the wild type band present.

Screening the RUN3 clones in this manner gave a pattern and size of the bands which were identical to those of the 'hit' from which they were derived, indicating that these 'run' clones had not excised the vector (figure 7.3).

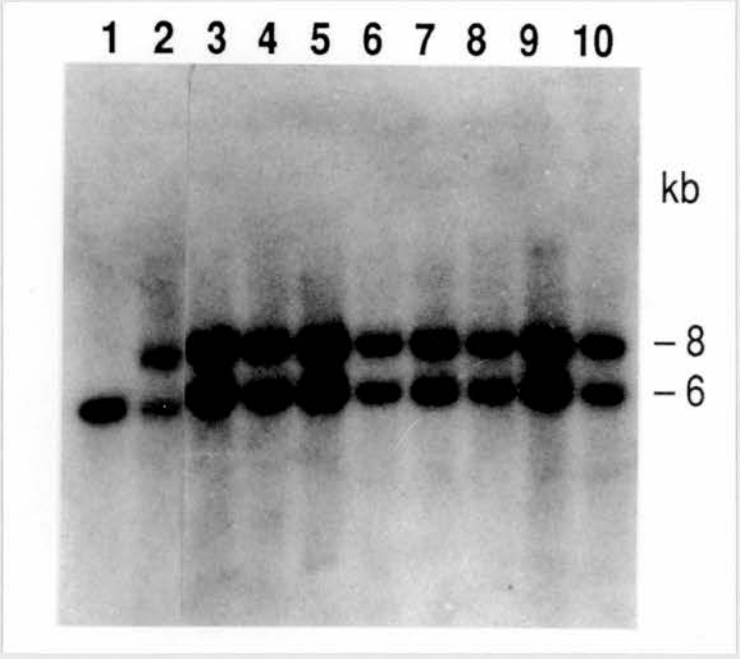


Figure 7.3 Screening RUN3 clones for the presence of the 'hit and run' targeting vector.
A southern blot of *Xba I-Sal I* double digests was probed with the 0.6EH genomic probe. Lane 1 shows the 6 kb wildtype band obtained with non-transfected CGR8 ES cells, and lane 2 the additional 8 kb band of the mutant allele created in 'hit' clones by integration of the targeting vector. Lanes 3 to 10 show the bands obtained from all RUN3 clones which are identical to that of the 'hit' clone from which they were derived.

7.3.3 Conclusions

This experiment was again unsuccessful in generating true 'run' clones which had excised the 'hit and run' targeting vector to become Gancyclovir resistant. The Gancyclovir selection was efficient at killing all cells expressing the Hygromycin-*tk* fusion gene, as demonstrated by the death of all Hygromycin resistant cells present on control plates when Gancyclovir selection was also imposed. Therefore the 'run' clones obtained in this experiment must be Gancyclovir resistant, and have survived Gancyclovir selection through some mechanism of loss of *tk* gene expression.

In contrast to the previous experiment, a higher number of Gancyclovir resistant clones were obtained from the plates in which the cells were plated at a higher density. However, the ten fold difference in the number of cells plated resulted in less than a ten fold difference in the number of clones obtained. The higher number of Gancyclovir resistant clones obtained at the higher plating density implies that at these densities, 'cross feeding' was not occurring. This is in contrast to the substantially lower numbers of Gancyclovir resistant clones obtained at the higher plating density in the previous experiment. Therefore it appears that the plating densities utilised in this experiment are sufficiently low to eliminate cross-feeding, thereby allowing the survival of cells which were not expressing *tk* gene. The generation of clones from plates which became confluent prior to addition of Gancyclovir however, does not fit with the 'cross feeding' theory. The majority of clones on these plates should be expressing the *tk* gene, and therefore the close cell contact when confluent should result in extensive cross feeding, and hence cell death. It is possible that in this situation, the period of cell growth which occurred prior to addition of Gancyclovir enabled the non-expressing cells to increase in numbers to such an extent that a dense colony was formed in which the majority of cells were protected from the toxic products of neighbouring cells. This would overcome the 'crossfeeding' effect experienced by cells when selection is imposed soon after seeding when the cells are at or near the vulnerable single cell stage of growth.

The number of Gancyclovir resistant clones did increase with the delay between plating and addition of the selection, however this could be a reflection of the increase in number of cells not expressing *tk* which were already present in the population of cells used to set up the 'run' selection. This possibility could be eliminated by growing the 'hit' clone in Hygromycin selection immediately prior to the Gancyclovir selection to ensure that all cells were initially expressing the Hygromycin-*tk* fusion gene. Any Gancyclovir resistant clones arising in the following 'run' selection should then be the result of recent loss of *tk* gene expression events, and an estimation of frequency could be determined.

The Gancyclovir resistant RUN3 clones which did survive to the analysis stage were found to still contain the integrated 'hit and run' targeting vector. As in RUN2, there did not appear to be any difference between the pattern and size of the bands obtained for the non-selected 'hit' clone and the RUN3 clones implying that there were no gross deletions or rearrangements present in the integrated vector. Therefore it appears that these clones have lost sensitivity to Gancyclovir by some mechanism which produces no obvious changes to the fusion gene detectable by the Southern blot analysis. As none of the cells were preselected for expression of the fusion gene prior to the Gancyclovir selection, it was not possible to determine whether this inactivation of *tk* occurred during the Gancyclovir selection, or if these cells were already present in the population of the 'hit' clone culture used to set up this experiment.

7.4 RUN4 'RUN' EXPERIMENT

A further 'run' experiment was set up this time using cells which had been preselected in Hygromycin selection. This experiment was set up in the same format as RUN3 so that the results could be compared and the effect of Hygromycin preselection determined. This should ensure that the starting population of cells were all expressing the fusion gene before the Gancyclovir selection was imposed, and

should eliminate any cells from the population which had already lost sensitivity to Gancyclovir by any mechanism.

7.4.1 Method

Cells from the same 1.10 'hit' clone cell culture used to set up RUN3 were grown under Hygromycin selection for 10 days, during which they reached a passage number of 28. The cells were seeded at the same two densities of 1 X 10^5, and 1 X 10^6 cells per 100 mm plate used in experiment RUN3, again upon PEF feeder layers. Gancyclovir selection was added sequentially to eight sets of plates on a daily basis as for RUN3, but in this experiment the period between plating and imposition of Gancyclovir selection also represented the period between Hygromycin selection and Gancyclovir selection. Therefore, cells which were seeded in the 'day 0' plates were switched from Hygromycin selection to Gancyclovir selection on the same day. Duplicate plates were seeded at each density for each day of Gancyclovir addition up until 'day 3', after which only one plate at each density was provided. Identical control plates to those of RUN3 were set up in which cells seeded at the density of 1 X 10^5 cells per 100 mm plate were maintained in Hygromycin selection in the absence of a PEF feeder layer, and Gancyclovir selection was added at daily intervals.

All plates were fed every three days with the appropriate medium and selection, and once visible, clones were counted, picked, and grown up for analysis.

7.4.2 Results

7.4.2.1 Clone numbers.

Gancyclovir resistant clones were obtained on average, eight days after Gancyclovir selection was imposed. As before, plates in which there was the longer delay before plating and addition of Gancyclovir selection, became confluent before the selection was imposed. No clones were visible on any of the control plates indicating that the Gancyclovir selection was working efficiently and killing all cells expressing the

Hygromycin-*tk* fusion gene. The numbers of clones obtained are detailed in table 7.5.

As in the previous experiment (section 7.3), the number of Gancyclovir resistant clones increased with the increased time between plating and imposition of the selection. However in this experiment, far fewer clones were obtained from the plates receiving the earlier additions of Gancyclovir than in RUN3, although by the end of the time course the numbers had reached comparable levels. At the earliest time points, there was very little difference between the number of clones obtained from plates seeded at the same density, although later on in the experiment the numbers of clones arising on plates seeded at the higher density became greater, and these plates reached a density that rendered them uncountable one day ahead of the plates seeded at the lower density.

Table 7.5 The numbers of Gancyclovir resistant clones obtained in experiment RUN4.

TIME OF GANC ADDITION (Days after plating)	SEEDING DENSITY (No. cells/plate x 10^6)	MEAN NUMBER OF CLONES PER PLATE
0	0.1	3
0	1.0	1
1	0.1	2.5
1	1.0	4
2	0.1	21.5
2	1.0	30
3	0.1	25.5
3	1.0	246
4	0.1	234
4	1.0	TMTC
5	0.1	236
5	1.0	TMTC
6	0.1	TMTC
6	1.0	TMTC
7	0.1	TMTC
7	1.0	TMTC

TMTC. Too many clones to count. Individual clones not distinguishable.

7.4.2.2 Screening RUN4 clones for vector excision.

A total of 78 RUN4 clones were picked and taken through to the analysis stage where they were screened for the presence of the 'hit and run' targeting vector by probing a Southern blot of *Bgl I* digests with the 0.6EH genomic probe (as described in 7.2.2.2). As in previous 'run' experiments this detected the 3.8 kb mutant band characteristic of the integrated hit and run targeting vector in all of the clones analysed, suggesting that the vector had not excised (figure 7.4).

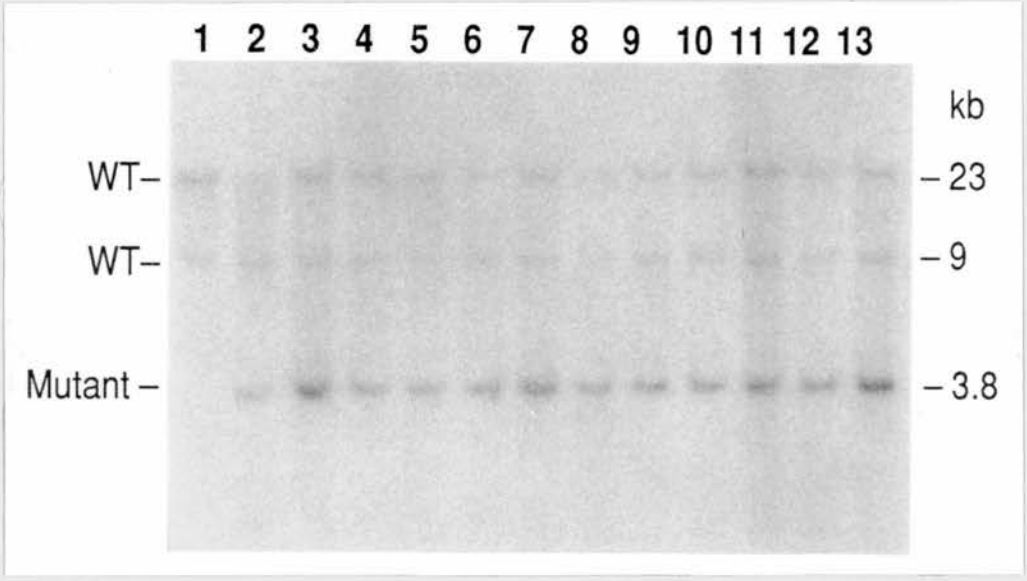


Figure 7.4 Screening RUN4 clones for excision of the 'hit and run' targeting vector.

A Southern blot of *Bgl I* digests probed with the 0.6EH genomic probe. Lane 1 shows the 23 kb and 9 kb wildtype bands of the non-transfected control, and lane 2 the additional 3.8 kb band of the mutant allele created in 'hit' clones by integration of the targeting vector. Lanes 3 to 13 show the bands obtained from all RUN4 clones which are identical to that of the 'hit' clone from which they were derived.

7.4.3 Conclusions

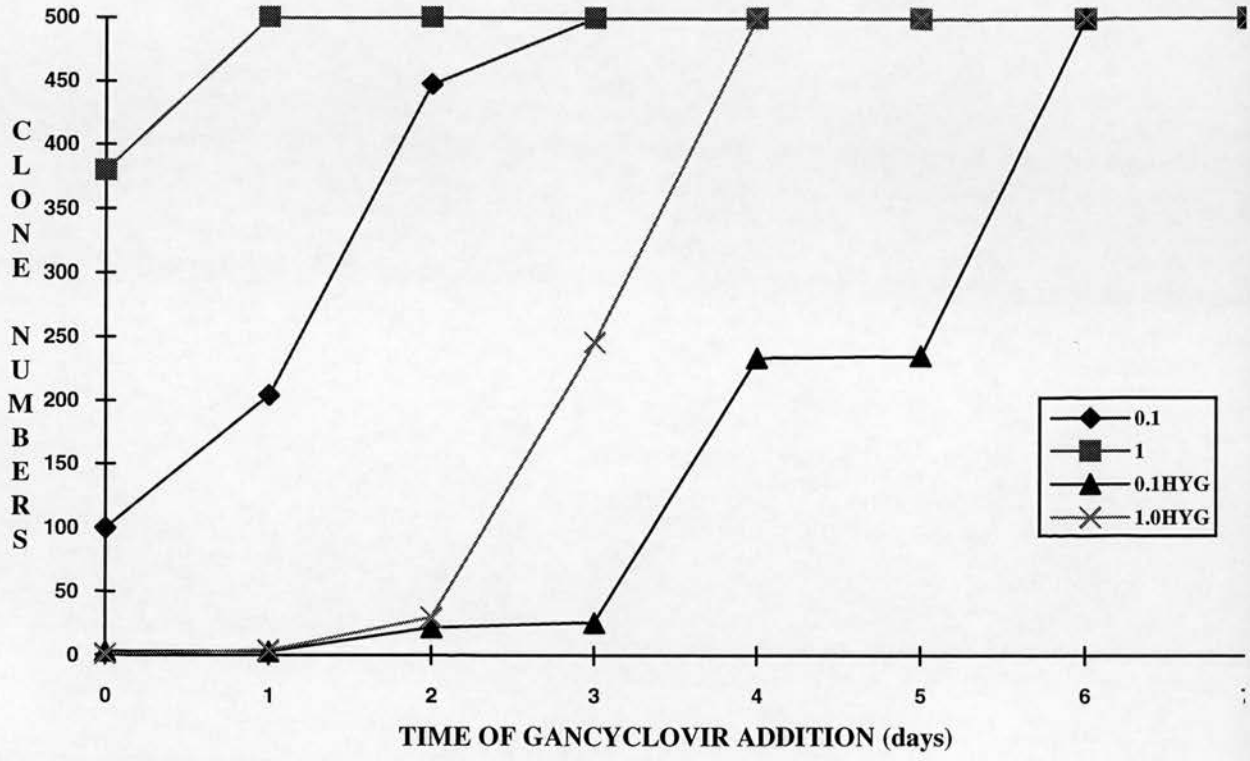
This 'run' experiment was consistent with previous experiments in its failure to generate 'run' clones in which the vector had excised, instead generating Gancyclovir resistant 'run' clones which were indistinguishable at the Southern blot level from the 'hit' clones from which they were derived. However the Hygromycin pre-treatment of the cells prior to their use in the 'run' stage, and the comparison with the similar RUN3 experiment which did not utilise this pre-treatment yielded some interesting data.

Preselection of 'hit' cells with Hygromycin prior to use for the 'run' selection should eliminate any cells which do not contain or express the fusion gene at a sufficient level to confer resistance to Hygromycin. Due to the nature of the Hygromycin-*tk* fusion gene, any mechanism which acts on the promoter to suppress Hygromycin expression should have the same effect upon *tk* expression, and therefore this preselection step should eliminate cells already present in the population which are Gancyclovir resistant through loss of the fusion gene expression. However, any cells which have specifically inactivated the *tk* part of the fusion protein e.g. a point mutation, without loss of Hygromycin resistance would not be eliminated by this preselection step. However, such cells would be identified on the control plates.

A comparison of the numbers of Gancyclovir resistant clones obtained in identical 'run' experiments utilising cells from the same 'hit' clone cell culture but which differed in their exposure to Hygromycin provided interesting data regarding the frequency of loss of Gancyclovir sensitivity (figure 7.5). The difference between the numbers of Gancyclovir resistant clones present at day 0 indicates that there was a significant number of Gancyclovir resistant clones already present in the 'hit' culture at the start of the experiment. The very low numbers of resistant clones obtained from the same culture which had been exposed to Hygromycin prior to Gancyclovir selection indicated that these clones were eliminated from the population by Hygromycin pre-treatment. This implies that the Gancyclovir resistance of these cells is likely to be due to an absence or dysfunction of the Hygromycin-*tk* fusion

gene product which also rendered them sensitive to Hygromycin. Therefore, many of the Gancyclovir resistant clones obtained from 'run' experiments in which the 'hit' clone had not been preselected with Hygromycin immediately prior to Gancyclovir selection are probably the progeny of previous events which have resulted in loss of Gancyclovir sensitivity. Loss of Gancyclovir sensitivity however, appears to be a relatively common event. Although Hygromycin pre-treatment appeared to eliminate virtually all Gancyclovir resistant cells from the starting population, their numbers rapidly increased with time until they reached a number which was comparable with those obtained from the culture which had not been exposed to Hygromycin. Therefore the Gancyclovir selection imposed to select for a 'run' event in which *tk* expression was lost through excision of the targeting vector, instead appeared to be selecting for a different mechanism of loss of *tk* gene activity which occurred in these 'hit' clones at a frequency of approximately 1×10^{-5} to 1×10^{-6} cells.

Figure 7.5 The number of Gancyclovir resistant clones obtained at two different plating densities in 'run' experiments RUN3 and RUN4.



A cut off level of 500 was taken as the level at which clone numbers were declared too many to count.
The legend label 0.1 refers to the numbers of clones obtained in experiment RUN3 from cells plated at a density of 0.1 X 10^6 cells/100 mm plate, and 1.0 the numbers of clones obtained in the same experiment from cells seeded at the higher density of 1.0 X 10^6 cells/100 mm plate. Clone numbers from cells preselected with Hygromycin in experiment RUN4 are labelled as 0.1 HYG for cells plated at 0.1 X 10^6 cells/100 mm plate, and 1.0 HYG for cells plated at a density of 1.0 X 10^6 cells/100 mm plate.

The presence of the 'hit and run' targeting vector in the 'run' clones demonstrated that although the pre-treatment with Hygromycin was effective in eliminating such clones from the starting population, the frequency of loss of sensitivity to Gancyclovir was such that by the end of the experiment a large number of such 'false positive' clones were generated. The complete absence of any 'run' clones in which the vector had excised indicated that whatever mechanism is responsible for the loss of *tk* gene activity, it was occurring at a much higher frequency than the desired 'run' event.

All experiments to date have utilised the same HR5 'hit' clone (1.10) to select for a 'run' event. The phenomenon in which 'run' clones derived from this 'hit' which are Gancyclovir resistant yet still retain the 'hit and run' targeting vector may be unique to this clone. Therefore this experiment should be repeated with another 'hit' clone to investigate this possibility.

7.5 RUN6 'RUN' EXPERIMENT

Selection for a 'run' event was conducted using a different 'hit' clone to the one used in previous experiments to determine if the unpredicted results were unique to the HR5 'hit' clone 1.10, or were a feature of all 'hit' clones.

7.5.1 Method

The 'hit' clone HR5 3.11 was grown on Hygromycin selection for seven days by which time it had reached a passage number of twenty two. Cells were seeded at a density of 1 X 10^5 cells per 100 mm plate onto PEF feeder layers, on sixteen plates in total. Gancyclovir was added to duplicate plates at daily intervals as before (RUN3 and RUN4), from day 0 to day 7. Control plates were included of 1 X 10^5 cells per 100 mm plate on Hygromycin selection, in the absence of PEF feeder layers as before (RUN3 and RUN4). Gancyclovir selection was added to the control plates at daily intervals as for selection for the 'run' event. The selection containing medium was replaced every two days, and once visible, clones were picked and grown for DNA analysis and frozen stocks.

7.5.2 Results

7.5.2.1 Clone numbers.

Gancyclovir resistant clones were visible on average, eight days after Gancyclovir selection was imposed. No clones were visible on the control plates, indicating that the Gancyclovir selection was working efficiently. As in previous experiments, those plates which received the later additions of Gancyclovir became confluent before this selection was imposed. The number of clones obtained are detailed in table 7.6. A

substantial number of clones was obtained on the day 0 plates, and the numbers of Gancyclovir resistant clones increased with increasing time between plating and addition of Gancyclovir selection, as in previous experiments.

Table 7.6 The number of Gancyclovir resistant clones obtained from experiment RUN6.

TIME OF GANC ADDITION (Days after plating)	TOTAL No. CLONES	No. CLONES PICKED
0	74	11
1	186	12
2	248	12
3	269	12
4	TMTC	0
5	TMTC	0
6	TMTC	0
7	TMTC	0
—	—	**47**

TMTC: Too many clones to count.

7.5.2.2 Screening RUN6 clones for excision of the 'hit and run' targeting vector.

A total of 24 RUN6 clones were taken through to the analysis stage where they were screened for the presence of the 'hit and run' targeting vector by probing a Southern blot of *Xba I-Sal I* double digests with the 0.6EH probe, a combination of digest and probe which had been used previously to detect targeted 'hit' clones (6.2.2.2). This revealed the same partially cut bands as those obtained when screening RUN2 clones by the same procedure (figure 7.6). These bands are indicative of the enzyme *Sal I* not cutting at the site present in the 'hit and run' targeting vector. However, the presence of the 16.5 kb *Xba I* fragment which represents the mutant allele of a targeted 'hit' clone, indicates that the targeting had not excised in these cells.

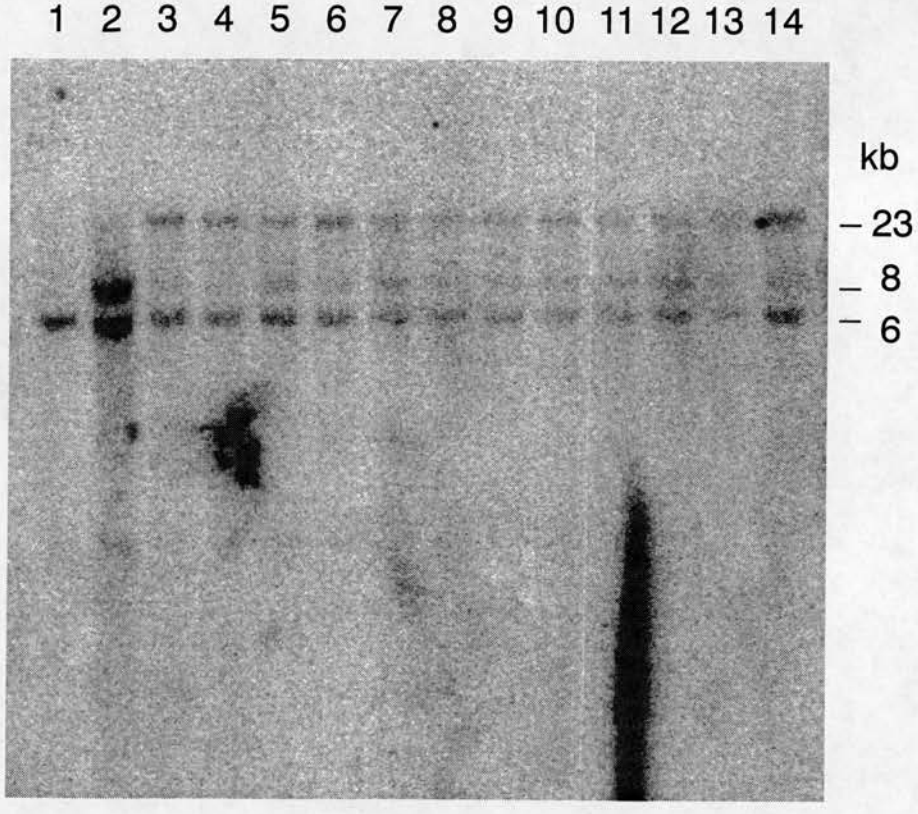


Figure 7.6 Screening RUN6 clones for the presence of the 'hit and run' targeting vector.

A Southern blot of *Xba I-Sal I* double digests was probed with the 0.6EH genomic probe. Lane 1 represents DNA from non-transfected CGR8 ES cells, and shows the 6 kb band of the wildtype allele. Lane 2 contains DNA from the 'hit' clone 1.10 and displays the 8 kb band of the mutant allele as well as the 6 kb wildtype band. Lanes 3 to 14 show results typical of all of the RUN6 clones. As well as the 6 kb wildtype band, 23 kb bands are present which represent the *Xba I* fragment of the target site containing the integrated 'hit and run' targeting vector, and the 8 kb band created by cutting the 23 kb *Xba I* fragment at the *Sal I* site present in the targeting vector. The intensity of the 8 kb band is faint, indicating that this *Sal I* site is not cutting to completion.

7.5 3 Conclusions

This experiment was unsuccessful in generating 'run' clones in which the 'hit and run' targeting vector had excised. The absence of clones on the control plates indicated that the Gancyclovir was working efficiently and killing those cells expressing the fusion gene. The Hygromycin preselection of the 'hit' clone should have ensured that all cells were expressing the Hygromycin-*tk* fusion gene at the start of the experiment. Therefore it appears that during Gancyclovir selection for the 'run' event, loss of *tk* gene activity occurred, allowing cells with the targeting vector to survive.

These results are in agreement with those obtained from previous experiments in which Gancyclovir selection was imposed upon a different 'hit' clone. This suggests that the phenomenon of clones surviving Gancyclovir selection with the Hygromycin-*tk* selection gene intact is not specific to one 'hit' clone.

A comparison of the numbers of Gancyclovir resistant clones obtained form 'hit' clone 3.11 and those obtained in the comparable 'run' experiment (RUN4) from 'hit' clone 1.10 are shown in figure 7.7. A higher number of Gancyclovir resistant clones are obtained from 'hit' 3.11 at the beginning of the time course than from 'hit' 1.10. Despite this, in both experiments the numbers clones rise slowly at approximately the same rate over the first three days, until day 4 when there is a rapid increase in clone numbers in both experiments. The higher number of clones obtained from day 0 plates in experiment RUN6 suggests that there were more Gancyclovir resistant clones present in the starting population of the 'hit' clone. However, Hygromycin selection should have removed all Gancyclovir resistant clones from the 'hit', and so this higher number of Gancyclovir resistant clones could be a consequence of a higher frequency of loss of *tk* activity in the 3.11 'hit' clone. The similar rates of increase in numbers of Gancyclovir resistant clones obtained from both 'hit' clones, suggests that a similar mechanism of Gancyclovir resistance is occurring in both 'hit' clones.

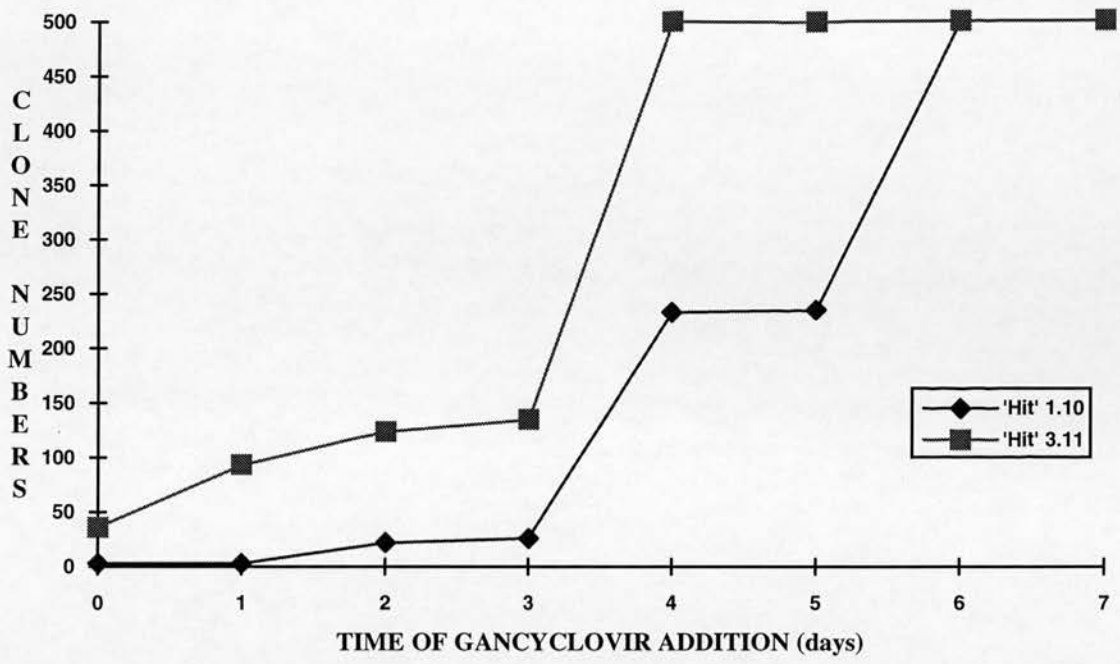


Figure 7.7 A comparison of the number of Gancyclovir resistant clones obtained from two different 'hit' clones under identical conditions.
The label 'Hit' 1.10 in the legend refers to the number of Gancyclovir resistant clones obtained from 'hit' clone 1.10 in experiment RUN4, and the label 'Hit' 3.11 refers to the number obtained from 'hit' clone 3.11 in experiment RUN6.

The results from this experiment utilising the 'hit' clone 3.11, and comparison with those utilising the 'hit' clone 1.10, demonstrate that the phenomenon of loss of Hygromycin-*tk* gene activity during Gancyclovir selection is common to both 'hit' clones. These 'hit' clones are the result of separate targeting events, as they arose on separate plates during a targeting experiment. Therefore, the properties they displayed are likely to be characteristic of pHTHRΔI506/7 'hit' clones in general, and indicate that loss of the Hygromycin-*tk* fusion gene activity during Gancyclovir selection is likely to be a common occurrence.

7.6 DISCUSSION

All 'run' experiments conducted so far had been completely unsuccessful in generating even a single clone in which the vector had excised. The Gancyclovir resistant clones arising from two different 'hit' clones appear to have retained the 'hit and run' targeting vector and the Hygromycin-*tk* fusion gene, and Southern blot analysis did not identify any differences between the mutant allele of the 'run' clones and that of the 'hit' from which they were derived. It appeared that some other mechanism was responsible for the loss of *tk* gene activity in these 'run' clones, which did not induce any obvious structural change to the mutant allele. These results do not exclude the possibility that excision of the vector is occurring in some cells, as this high level of 'false positive' Gancyclovir resistant clones could be obscuring such an event if it were occurring at a low frequency.

The observation that clones can survive a negative selection with the negative selection gene of a vector seemingly intact, has also been reported by others when targeting different loci in ES cells. Wu *et al.* (1994), used the 'hit and run' targeting strategy to introduce base substitutions into the murine type I collagen gene. They found that 32% of 'run' clones which were resistant to FIAU (an alternative selection against *tk* gene expression), had still retained the targeting vector. A similar figure of 38% was reported by Ramirez-Solis *et al.* (1993) when targeting the *Hoxb-4* locus with this technique, and the lower figure of 22% of cells surviving the negative selection with the intact targeting vector was reported by Valancius and Smithies (1991) when introducing mutations into the *hprt* gene using a version of the 'hit and run' strategy which they called the 'In-Out' targeting procedure. Use of alternative strategies for the introduction of subtle mutations into genes, which also utilised a negative selection step have encountered this problem to varying degrees. Askew *et al.* (1993) found that 92% of Gancyclovir resistant clones obtained from the second step of a variation of the double replacement targeting strategy also contained the intact replacement vector. Analysis of the *tk* gene within the replacement vector detected no gross rearrangements. The introduction of a small deletion into the murine type I collagen gene by Wu *et al.* (1994) this time using a double replacement

strategy, also found that 40% of FIAU resistant cells contained the *tk* gene. Therefore although the proportion of resistant clones varied, these observations of clones surviving negative selections with the negative selection genes intact suggests this phenomenon was not unique to the *Cftr* locus or this targeting vector. They also show that loss of sensitivity to a negative selection was not a feature peculiar to the use of Gancyclovir to select against *tk* gene expression as a negative selection system, as some of these experiments utilised FIAU to select against Gancyclovir expression, and others utilised selection for and against the *hprt* gene as their selection system.

This results however, are in contrast to comparable experiments carried out by others in our own lab to introduce the ΔF508 mutation into *Cftr* using a similar (but not identical) 'hit and run' targeting vector. In these experiments, negative selection generated many clones in which the ΔF508 vector was absent, indicating that they had excised the vector and reverted to a wildtype genotype. The main difference between the ΔF508 'hit and run' targeting vector and the ΔI506/7 'hit and run' targeting vector was that the positive and negative selections were expressed from two genes under the control of two separate promoters. Also, the ΔF508 'hit and run' targeting vector employed Neomycin as the positive selection rather than the Hygromycin selection utilised by the ΔI506/7 'hit and run' targeting vector. The difference in the nature of Gancyclovir resistant clones generated by the two vectors might therefore be a result of the different selection cassettes, and this possibility should be investigated before any further 'run' experiments were conducted.

The conclusions from these experiments can be summarised as follows:

1. Cells were surviving Gancyclovir selection with the Hygromycin-*tk* fusion gene intact.
2. Southern blot could not distinguish any differences between mutant allele of 'hit' and that of 'run' clones.
3. The Hygromycin-*tk* fusion gene of the 'hit' clone from which 'run' clones were derived was functional.

4. Gancyclovir selection was shown to be effective.
5. Loss of *tk* gene activity therefore occurred during 'run' stage.
6. Control plates indicated that loss of Gancyclovir sensitivity was a result of loss of both Hygromycin and *tk* activity of fusion gene product.

These experiments and the observations of others, imply that some mechanism other than vector excision was responsible for loss of *tk* gene activity during Gancyclovir selection. The high frequency at which this inactivation occurred and the large number of resistant clones generated could be masking any resistant clones which arose through the desired mechanism of vector excision. Therefore it was important before proceeding any further, to identify the mechanism by which cells containing the 'hit and run' targeting vector were able to inactivate *tk* gene expression and survive Gancyclovir selection.

CHAPTER 8

INVESTIGATION OF THE GANCYCLOVIR RESISTANCE OF 'RUN' CLONES

8.1 INTRODUCTION

Previous experiments in which Gancyclovir selection was imposed upon targeted 'hit' clones to select for excision of the 'hit and run' targeting vector had produced clones which were Gancyclovir resistant without the vector excision event. All of the clones examined had retained the targeting vector which implies that this high background could be masking the presence of any clones which had lost *tk* expression through the desired mechanism of vector excision.

Further analysis was carried out on Gancyclovir resistant clones arising from one of these 'run' experiments to identify the mechanism by which the *tk* gene had become inactivated so that possible ways of overcoming this might be developed.

8.2 INVESTIGATION OF THE SELECTION RESISTANT PHENOTYPE OF RUN4 CLONES

In order to determine if both the Hygromycin and the *tk* components of the fusion gene had been inactivated in the RUN4 clones, a number were tested for Hygromycin and Gancyclovir resistance.

8.2.1 Method

Twenty five 'run' clones, plus the 'hit' clone from which they were derived, were each seeded at a density of 4 X 10^4 cells into four wells of a twenty four well plate. Twenty four hours after seeding, Hygromycin selection was imposed at the usual concentration of 150 μg/ml upon one of the four wells of each clone. Gancyclovir selection was imposed onto another well, and Hygromycin and Gancyclovir selection combined onto a third, leaving the fourth well of each clone as a non-selected control. The wells were fed every two days, and the growth of each clone noted after two weeks.

8.2.2 Results

The resistance of each clone to either selection varied widely, with different combinations of sensitivity and resistance observed (table 8.1). A large number of clones were resistant to Hygromycin selection and sensitive to Gancyclovir selection, a phenotype consistent with a functional Hygromycin-*tk* fusion gene. However, most of these clones also produced a variable background of Gancyclovir resistant clones. A few clones appeared to be sensitive to Hygromycin and resistant to Gancyclovir selection, a phenotype consistent with inactivation of the selection fusion gene, and also gave rise to a varying background, this time of Hygromycin resistant cells. Some clones gave a phenotype which was completely at odds with the presence of the integrated 'hit and run' vector, that is resistance to both selections.

Table 8.1 The selection resistant phenotype of RUN4 clones.

PHENOTYPE	CLONE
GROUP I	18.3
Hygromycin resistant	12.6
Gancyclovir sensitive	
No background resistance	
GROUP II	Hit 1.10
Hygromycin resistant	18.1
Gancyclovir sensitive	11.2
Few Gancyclovir^R colonies	8.3
	9.7
	9.8
	10.1
	13.5
	16.2
	12.1
	21.1
GROUP III	11.4
Hygromycin resistant	11.7
Gancyclovir sensitive	14.4
High Gancyclovir^R	14.7
GROUP IV	1.2
Hygromycin resistant	9.4
Gancyclovir resistant	11.1
	13.7
GROUP V	
Hygromycin sensitive	11.8
Gancyclovir resistant	
GROUP VI	
Hygromycin sensitive	6.3
Gancyclovir resistant	
Background Hygromycin^R	
GROUP VII	7.1
Hygromycin sensitive	6.2
Gancyclovir resistant	
High Hygromycin^R	

8.2.3 Conclusions

The presence of the 'hit and run' targeting vector with a functional Hygromycin-*tk* fusion gene should confer Hygromycin resistance and Gancyclovir sensitivity on the RUN 4 clones. This phenotype however, was only displayed by some of the clones. If the fusion gene had become inactivated, the reverse phenotype of Hygromycin sensitivity and Gancyclovir resistance should then be observed. Although these two phenotypes were seen in a number of clones, they did not account for all of the phenotypes observed. Some clones displayed the bizarre phenotype of resistance to both selections (Hygromycin but not *tk* gene activity), which is at odds with Hygromycin resistance and Gancyclovir sensitivity being conferred by the same fusion protein. In addition, clones which appeared to be Gancyclovir sensitive also displayed various levels of background resistance, as did Hygromycin sensitive clones. It is also interesting to note that no clones were found to be sensitive to both selections (i.e. exhibiting *tk* gene activity only).

The variety of phenotypes displayed by the 'run' clones were unexpected as it had been assumed that they had all survived Gancyclovir selection through the same mechanism of inactivation of the Hygromycin-*tk* fusion gene. It was surprising that a large number of clones were sensitive to Gancyclovir as these clones should not have survived the 'run' Gancyclovir selection. This might suggest that the Gancyclovir selection was not effective, however large scale cell death occurred when Gancyclovir was imposed on the 'hit' clone during the 'run' step, giving rise to clones rather than confluent plates. In addition, the absence of clones on the control plates on which Hygromycin and Gancyclovir selection were imposed together implied that the selection was working efficiently as clones would have arisen on these plates if the Gancyclovir selection had not killed cells which were expressing the fusion gene. This phenotype of Hygromycin resistance and Gancyclovir sensitivity is what is expected for clones which still contain and express the intact 'hit and run' targeting vector. Therefore, it is difficult to account for how these Gancyclovir sensitive clones survived the Gancyclovir selection as they appear to be expressing the selection fusion gene. The various levels of background Gancyclovir

resistant cells obtained from the Gancyclovir sensitive clones however, suggests that some cells in this clone population did not produce the Thymidine kinase fusion protein.

The Hygromycin sensitive and Gancyclovir resistant phenotype displayed by some clones was the phenotype which would be expected for cells in which the fusion gene was not expressed or lost. The presence of the 'hit and run' targeting vector in these clones indicated that in these cells, the selection fusion gene had become inactivated. As cells were preselected for those which were expressing the fusion gene as indicated by Hygromycin resistance prior to setting up the 'run', this inactivation of the fusion gene must have occurred during selection for the 'run' event. The various levels of Hygromycin resistant cells arising from these clones indicates that the fusion gene was not inactivated in all of the cells from these clone populations.

The observation that some clones were resistant to both selections is inconsistent with the expression of both selections from a single fusion gene. Resistance to Hygromycin is observed in cells which are expressing the Hygromycin-*tk* fusion gene, and therefore these cells should also be sensitive to Gancyclovir. Any mechanism which has inactivated the *tk* gene expression of these clones must have not affected Hygromycin expression. Loss of *tk* gene activity alone by the Hygromycin-*tk* fusion protein could arise as a result of small changes to the *tk* gene, such as point mutations which would not disrupt its Hygromycin resistant properties. Alternatively, the level of expression of the fusion gene could differentially affect the Hygromycin resistance and *tk* sensitivity. A low level of expression of the fusion gene might be sufficient to confer resistance to Hygromycin whereas a much higher level of expression might be required to bestow sensitivity to Gancyclovir. However, both of these mechanisms of selective *tk* gene inactivation can be discounted as they would both give rise to clones on the control plates on which both Hygromycin and Gancyclovir had been imposed simultaneously. The complete absence of any clones in combined selection indicates that a single cell within a clone population was capable of displaying either a discrete Hygromycin resistant and Gancyclovir

sensitive, or a Hygromycin sensitive and Gancyclovir resistant phenotype only, consistent with complete expression or inactivation of the Hygromycin-*tk* fusion gene. This implies that the resistance to both selections displayed by some clones must be attributable to a mixed population of expressing and non-expressing cells. Hygromycin would select for those cells in the population which are expressing the fusion gene, and Gancyclovir for those in which such expression is absent. The background of resistant cells observed with Hygromycin sensitive clones could therefore represent a minority of expressing cells in a mainly non-expressing population, whereas the Gancyclovir resistant background of a Gancyclovir sensitive clone might represent a minority of expressing cells in a clone population.

Therefore it appears that the 'run' clones were mixed populations with a variable proportion of cells which did not express the Hygromycin-*tk* fusion gene, as observed by their selection resistant phenotype. A large number of 'run' clones exhibited a Gancyclovir sensitive phenotype which was inconsistent with their original generation from Gancyclovir selection of the 'hit' clone. The mechanism by which the Hygromycin-*tk* fusion gene had become inactivated was unknown.

8.3 FURTHER SOUTHERN BLOT ANALYSIS OF RUN4 CLONES.

The RUN4 clones which had been tested by addition of the different selection regimes were analysed further by Southern blotting.

8.3.1 Method

A Southern blot utilising the digest and probe combination which had been used previously to detect 'hit' clones (chapter 6) was conducted to determine if there were any differences between the 'run' clones and the 'hit' from which they were derived, which might account for their altered selection resistance. For this, RUN4 clone DNA was digested with *Xba I* and *Sal I*, and probed with the 0.6EH probe.

8.3.2 Results

This analysis yielded complex pattern of bands which appeared to represent a failure of the restriction enzyme *Sal I* to cut the DNA at the site present in the 'hit and run' targeting vector (figure 8.1). The degree of cutting of this enzyme appeared to vary between the RUN4 clones, but consistently cut DNA from the non-selected 'hit' clone to completion. On close examination, the degree of cutting at the *Sal I* site in the targeting vector appeared to correlate with the selection resistant phenotype of the RUN4 clones, and is presented in table 8.2.

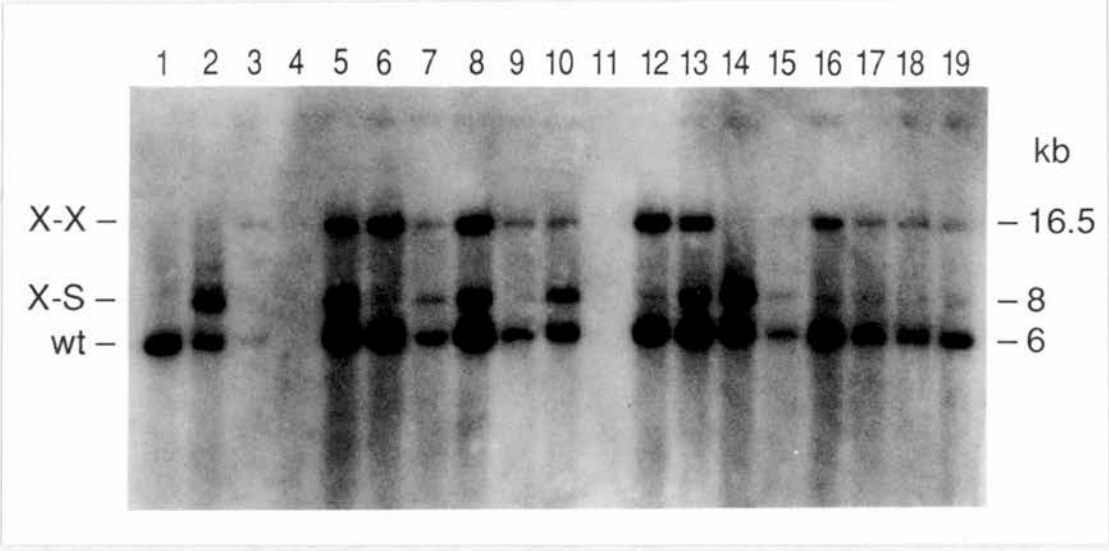


Figure 8.1 Further Southern blot analysis of 'RUN4' clones.
Clone DNA was digested with *Xba I* and *Sal I* restriction enzymes and probed with the 0.6EH genomic probe. Lane 1 is the non-transfected control displaying the wildtype 6 kb *Xba I* fragment generated at a wildtype locus (indicated by wt). Lane 2 is the 'hit' clone which has not been subject to Gancyclovir selection. This displays the wildtype 6 kb band as well as the 8 kb mutant band generated when the 16.5 kb *Xba I* fragment created by insertion of the targeting vector into the target site is cut at the *Sal I* site present in the targeting vector (labelled X-S). Lanes 3 to 19 represent RUN4 clone DNA which all display the 6 kb wildtype band. However the intensity of the 8 kb mutant band (X-S) varies between each clone, with a predominance of the 16.5 kb band (X-X) representing the increased *Xba I* fragment created by insertion of the wildtype allele which has not cut at the *Sal I* site present in the vector. For identification on table 8.2, the RUN4 clones in each lane are as follows: Lane (3) 1.2, (4) 6.2, (5) 8.3, (6) 9.4, (7) 9.7, (8) 9.8, (9) 11.1, (10) 11.2, (11) 11.4, (12) 11.7, (13) 12.1, (14) 12.6, (15) 13.5, (16) 13.7, (17) 14.4, (18) 14.7, (19) 16.2.

Table 8.2 The selection resistant phenotype of RUN4 clones and their correlation with the degree of cutting at the *Sal I* site of the 'hit and run' targeting vector.

PHENOTYPE	CLONE	MUTANT *Xba.1* AND *Sal.1* BANDS OBSERVED	DEGREE OF *Sal I* DIGESTION (%)
GROUP I	18.3	*Xba-Sal*	100
Hygromycin resistant	12.6	*Xba-Sal*	100
Gancyclovir sensitive			
No background resistance			
GROUP II	Hit 1.10	*Xba-Sal*	100
Hygromycin resistant	18.1	*Xba-Sal*	100
Gancyclovir sensitive	11.2	*Xba-Xba + Xba-Sal*	70
Few Gancyclovir^R colonies	8.3	*Xba-Xba + Xba-Sal*	50
	9.7	*Xba-Xba + Xba-Sal*	50
	9.8	*Xba-Xba + Xba-Sal*	50
	10.1	*Xba-Xba + Xba-Sal*	50
	13.5	*Xba-Xba + Xba-Sal*	50
	16.2	*Xba-Xba + Xba-Sal*	50
	12.1	*Xba-Xba + Xba-Sal*	20
	21.1	*Xba-Xba + Xba-Sal*	20
GROUP III	11.4	*Xba-Xba + faint Xba-Sal*	10
Hygromycin resistant	11.7	*Xba-Xba + faint Xba-Sal*	10
Gancyclovir sensitive	14.4	*Xba-Xba + faint Xba-Sal*	10
High Gancyclovir^R	14.7	*Xba-Xba + faint Xba-Sal*	10
GROUP IV	1.2	*Xba-Xba*	0
Hygromycin resistant	9.4	*Xba-Xba*	0
Gancyclovir resistant	11.1	*Xba-Xba*	0
	13.7	*Xba-Xba*	0
GROUP V			
Hygromycin sensitive	11.8	*Xba-Xba*	0
Gancyclovir resistant			
GROUP VI			
Hygromycin sensitive	6.3	*Xba-Xba*	0
Gancyclovir resistant			
Background Hygromycin^R			
GROUP VII	7.1	*Xba-Xba + faint Xba-Sal*	10
Hygromycin sensitive	6.2	*Xba-Xba + faint Xba-Sal*	10
Gancyclovir resistant			
High Hygromycin^R			

Xba-Xba represents the 16.5 kb *Xba* I fragment generated when the *Sal I* site present in the vector does not cut. *Xba-Sal* represents the 8 kb fragment generated by cutting the 16.5 kb *Xba-Xba* fragment at the *Sal I* site introduced by the vector.

8.3.3 Conclusions.

There is a strong correlation between the degree of cutting at the *Sal I* site in the 'hit and run' targeting vector with the selection resistant phenotype of the RUN4 clones. Clones whose phenotype suggested that the fusion gene had been inactivated, exhibited bands on the Southern blot indicative of a complete failure of the restriction enzyme *Sal I* to cut at the restriction site in the 'hit and run' targeting vector. In contrast, clones whose phenotype appeared to indicate that the fusion gene was active, gave bands indicative of *Sal I* digestion at the restriction site in the vector, although the degree of cutting at this site decreased in clones with increasing levels of background Gancyclovir resistance. Therefore whatever mechanism was causing loss of the Hygromycin-*tk* fusion gene activity, also appeared to be preventing the restriction enzyme *Sal I* from cutting at the site in the 'hit and run' targeting vector.

The *Sal I* restriction site of the 'hit and run' targeting vector is situated at the 3' end of the Hygromycin-*tk* fusion gene, in the sequence specifying the *tk* gene and hence Gancyclovir resistance. It is compelling that cells which have appeared to survive Gancyclovir selection through inactivation of the *tk* gene, fail to cut at a *Sal I* restriction site located in this gene. Analysis showed that this enzyme does however cut to completion at the *Sal I* site present in the targeting vector of cells from the targeted 'hit' clone which had not been subjected to Gancyclovir selection. This suggests that the *Sal I* restriction site was intact in the 'hit' clones used for the 'run' selection, and that whatever event was responsible for preventing the enzyme from cutting at this site in the 'run' clones probably occurred during the 'run' stage. This also implies that whatever mechanism was responsible for inactivation of the Hygromycin-*tk* fusion gene, also prevented the *Sal* I restriction enzyme from cutting at the site in the selection gene.

The restriction enzyme *Sal I* is known to be sensitive to methylation of CpG sequences, which it is unable to cut (Arrand *et al.* 1978). Methylation of CpG sequences is also a recognised mechanism of gene inactivation which does not

mutate the gene in any way (see discussion 8.6). Therefore it is possible that the mechanism by which the RUN4 clones were able to survive Gancyclovir selection was by methylation-induced suppression of the Hygromycin-*tk* gene expression. As methylation does not cause any physical alteration or mutation to the DNA sequences, methylation-induced gene suppression is a reversible effect. The background Hygromycin or Gancyclovir resistance seen with most of the clones could be due to a minority of cells in the population which have a different methylation status, and hence different expression of the Hygromycin-*tk* fusion gene. The degree of cutting at the *Sal I* site in the targeting vector could be a reflection of the methylation status of the clone cell population, and also an indicator of fusion gene expression. If this were the case, it would suggest that clones which exhibited complete inhibition of cutting at the vector *Sal I* site, and which did not exhibit fusion gene activity, were Gancyclovir resistant due to methylation-induced suppression of the fusion gene expression. Conversely, clones which did express the fusion gene and did cut to quite a high degree at the *Sal I* site, would not be methylated. In addition, the background Gancyclovir resistant cells seen with Gancyclovir sensitive clones could represent methylated cells in a largely non-methylated population, and perhaps the original methylation status of the clone which had allowed it to survive the Gancyclovir selection of the 'run' step. The Hygromycin resistant background of the Hygromycin sensitive clones would in turn represent a non-methylated sub-population of cells in a largely methylated population. The clones which were resistant to both selections could therefore also represent methylated and non-methylated subpopulations which were either expressing or not expressing the fusion gene, and were therefore resistant to both selections.

8.4 CORRELATION OF *Sal I* DIGESTION WITH FUSION GENE EXPRESSION.

It has been proposed that clones which are Gancyclovir resistant despite containing the 'hit and run' targeting vector, have inactivated the selection genes through

methylation. These clones have also been suggested to be mixed populations with varying levels of non-methylated expressing cells and methylated non-expressing cells which was reflected by the degree of cutting of the restriction enzyme *Sal I* at the site in the *tk* gene sequences of the fusion gene.

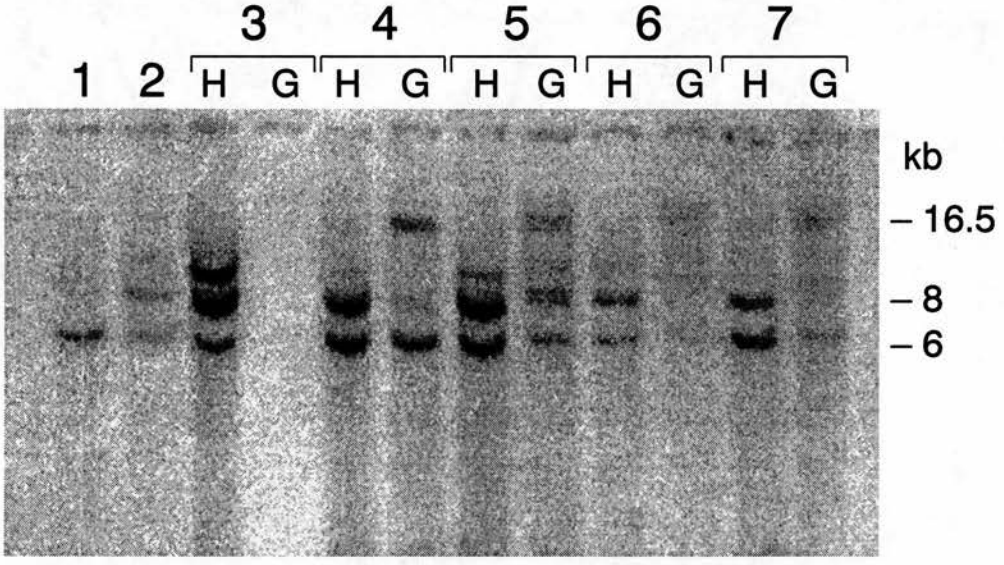
An experiment was conducted to ascertain whether the degree of cutting at the *Sal I* restriction site in the 'hit and run' targeting vector did correlate with expression of the fusion gene, as demonstrated by Hygromycin resistance and Gancyclovir sensitivity.

8.4.1 Method

RUN4 clones were seeded into two wells of a twenty four well plate as before (8.2.1), and Hygromycin selection was imposed upon one well, and Gancyclovir selection upon the other. Any cells which were resistant to the selection, including those demonstrating just a low background resistance, were grown until confluent. Once confluent, the cells were harvested and DNA extracted. A Southern blot was then performed on *Xba I-Sal I* digested DNA, which was probed with the 0.6EH probe as before (8.3.1).

8.4.2 Results

The bands obtained indicated that DNA from all of the clones which had been grown on Hygromycin cut almost completely at the *Sal I* site in the targeting vector to generate the expected 8 kb band (figure 8.2). However, cells from the same clones which had been grown on Gancyclovir selection exhibited very little cutting at the this site, resulting mainly in a 16.5 kb *Xba* I mutant band. This result was obtained for all of the clones which were tested, regardless of the selection resistant phenotype that they had previously displayed (table 8.2).


3
4
5
6
7
1
2
H
G
H
G
H
G
H
G
H
G
kb
– 16.5
– 8
– 6

8.4.3 Conclusions

Cells which are expressing the fusion gene will be Hygromycin resistant and will be selected for in Hygromycin selection whereas cells not expressing the fusion gene will be selected for in Gancyclovir selection. Therefore, the observation that selection of RUN4 clones in Hygromycin selection also induces cutting at the *Sal I* site in the 'hit and run' targeting vector, whereas selection in Gancyclovir represses cutting at this site in the same clone suggests there is a definite link between fusion gene activity and cutting at the *Sal I* site of the targeting vector.

It is also interesting that the same effect was seen in all clones tested despite the differing phenotypes displayed previously (table 8.1). When cells which gave rise to the Gancyclovir background resistance of Gancyclovir sensitive clones were analysed, they displayed an inhibition of *Sal* I digestion at the site in the vector. Conversely, the cells which gave the Hygromycin resistant background of Hygromycin sensitive clones showed almost complete cutting at the *Sal* I site. Therefore it appears that the 'run' clones are mixed populations with respect to activity of the fusion gene, and that different subpopulations of cells can be selected for by Hygromycin or Gancyclovir selection. The differential cutting at the *Sal I* site in the targeting vector of the same clone when grown on the two different selections suggests that this difference in expression could be due to methylation-induced gene suppression which has inactivated the fusion gene and also inhibited *Sal I* cutting in this region.

8.5 AN INVESTIGATION INTO THE METHYLATION STATUS OF THE RUN4 CLONES

The results from the previous experiments have shown that inactivation of the Hygromycin-*tk* fusion gene is linked to a failure of the *Sal I* restriction enzyme to cut at a site in this gene. *Sal I* is an enzyme which is known to cut only non-methylated DNA and it is therefore possible that methylation within the fusion gene is

inactivating the gene expression. The methylation status of the RUN4 clones was investigated by the use of other methylation sensitive enzymes.

8.5.1 Method

RUN4 clone DNA was digested separately with two different enzymes, *Hpa II* and *Msp I*. These enzymes both cut at the same site however, *Hpa II* does not cut methylated DNA. Therefore, any differences between the bands obtained from these digests on a Southern blot, identifies regions in which *Hpa II* has not cut, indicating that the DNA is methylated. In order to detect differences in methylation in different regions of the targeted *Cftr* allele, the Southern blot of these digests was hybridised with probes from different regions of the target site:

(I) A probe consisting of the whole of the Hygromycin-*tk* fusion gene was used to identify methylated regions within the sequences encoding the selection resistance.

(II) The filter of the *Hpa II* and *Msp I* digests was stripped and reprobed with the 0.6EH genomic probe from 3' region external to the integrated targeting vector. This should detect any methylation present in the 3' sequences flanking the target site.

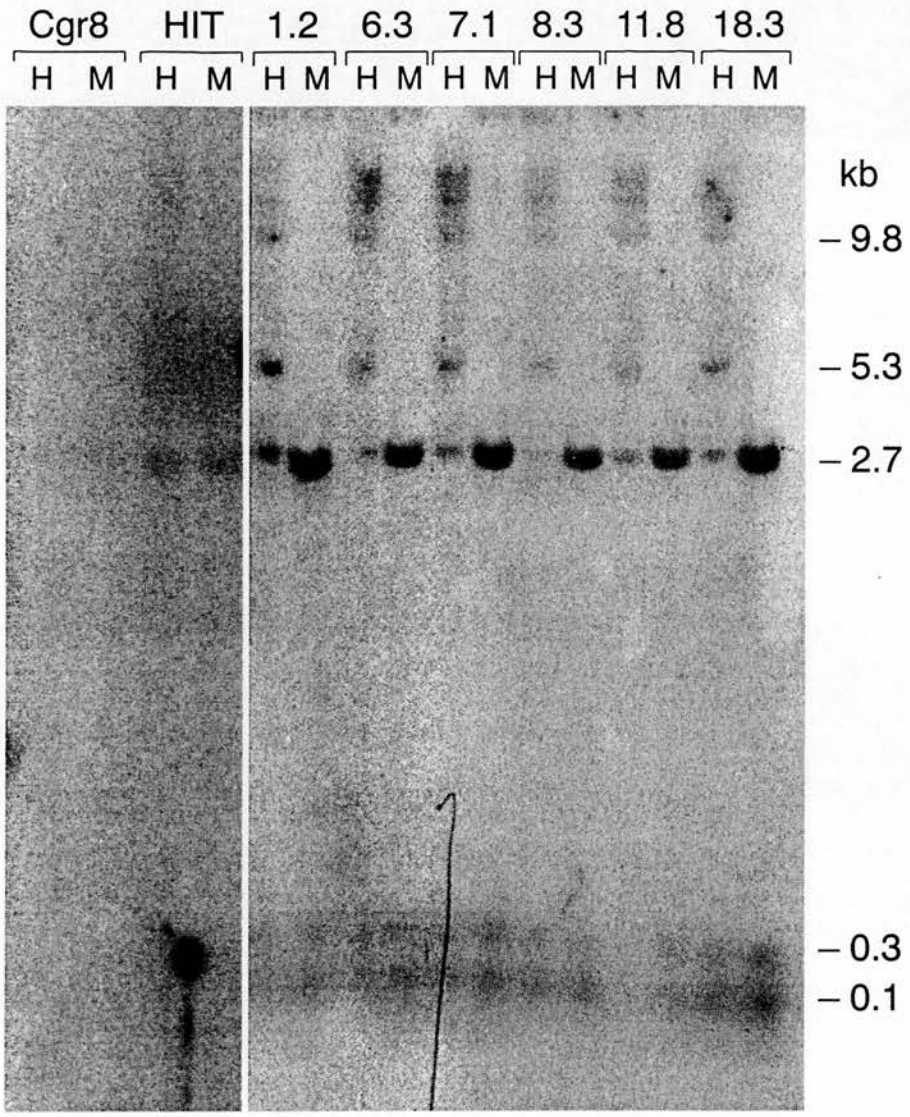
(III) The same filter was also probed with the 1.3XH probe. This should identify any methylation differences in the region of *Cftr* which is 5' to the integrated vector.

8.5.2 Results

A difference in the degree of cutting of the two enzymes was seen with all three probes. In all cases, the RUN4 clones appeared to contain regions of methylated DNA.

(I) Hygromycin-tk Probe

This probe detected three bands of 2.7, 0.3, and 0.1 kb from DNA which had been digested with the *Msp I* restriction enzyme (figure 8.3). However the 2.7 kb band was much fainter or absent from DNA digested with the methylation sensitive enzyme *Hpa II*, and in addition larger bands of 5.3 and 9.8 kb were also visible. The absence of any differences between the *Hpa II* and *Msp I* bands in the 'hit' clone suggests that this methylation is unique to the fusion gene of the Gancyclovir selected 'run' clones. The specificity of the probe for the Hygromycin-*tk* fusion gene sequences was demonstrated by the absence of bands in the non-transfected control.


Cgr8
HIT
1.2
6.3
7.1
8.3
11.8
18.3
H M H M H M H M H M H M H M H M
kb
– 9.8
– 5.3
– 2.7
– 0.3
– 0.1

(II) 3' 0.6EH Probe

The bands obtained with this probe also differed between the two digests, with a 1.6 kb band present in the *Msp I* digested DNA being absent from the same DNA digested with *Hpa II* (figure 8.4). Instead, larger bands of 5.3 and 6.6 kb were visible in the *Hpa II* digests. This indicates that the methylated DNA sequences of the RUN4 clones were not confined to the Hygromycin-*tk* fusion gene, and that the genomic region flanking the 3' end of the integrated 'hit and run' targeting vector was also methylated. The 1.6 kb band was also absent in the 'hit' clone and the non-transfected control, indicating that this region is also methylated in cells which have not undergone Gancyclovir selection.

(III) 1.3XH Probe

This probe detected an 8 kb band in the *Msp I* digests which was absent in the *Hpa II* digests indicating that the 5' region of *Cftr* was also methylated (figure 8.5). This altered band pattern was again seen in the 'hit' clone as well as the non-transfected control indicating that this region is normally methylated in cells which have not been selected in Gancyclovir.

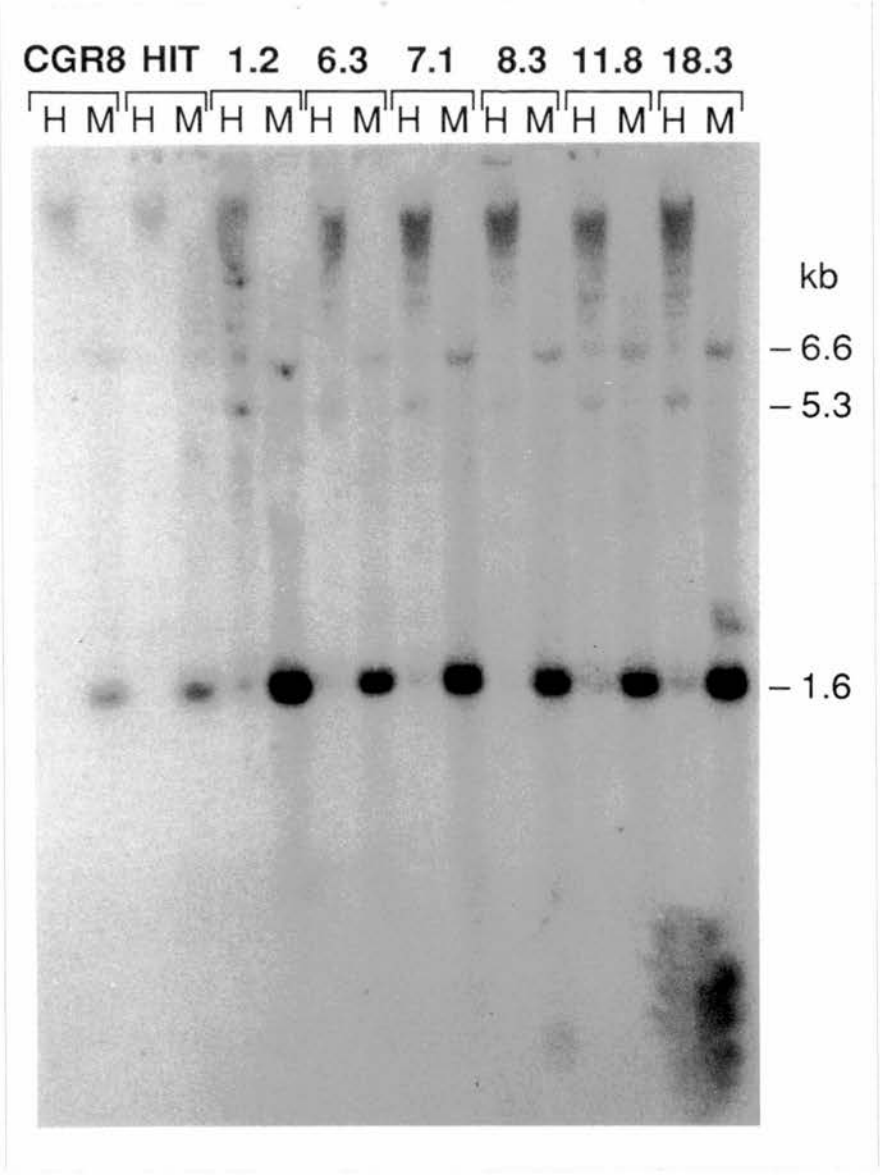


Figure 8.4 Analysis of the methylation status of the 3' sequences flanking the target site by Southern blot utilising methylation-sensitive restriction enzymes.
A Southern blot of DNA digested with restriction enzymes *Hpa II* or *Msp I* was reprobed with the 0.6EH probe. All of the lanes denoted by H contain DNA which has been digested with the restriction enzyme *Hpa II*, and the lanes labelled M, DNA which had been digested with *Msp I*. The lanes labelled CGR8 contain DNA from a non-transfected control, and those labelled 'hit' DNA from the 'hit' clone from which the 'run' clones were derived. The remaining pairs of lanes contain DNA from RUN4 clones.

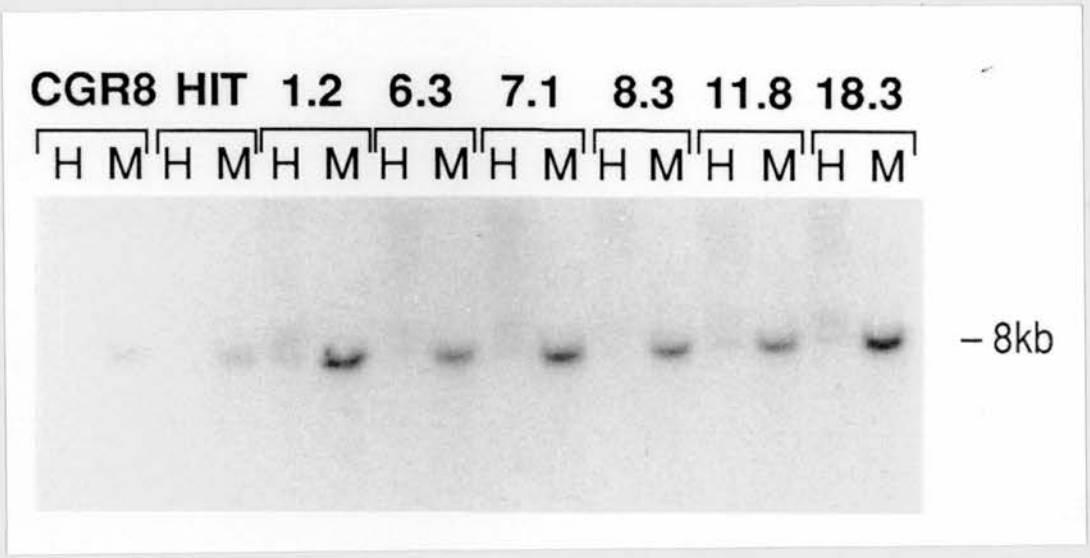


Figure 8.5 Analysis of the methylation status of the 5' sequences flanking the integrated targeting vector of RUN4 clones by Southern blot utilising methylation-sensitive restriction enzymes.
A Southern blot of DNA digested with restriction enzymes *Hpa II* or *Msp I* was probed with the 1.3XH genomic probe. All of the lanes denoted by H contain DNA which has been digested with the restriction enzyme *Hpa II*, and the lanes labelled M, DNA which had been digested with *Msp I*. The lanes labelled CGR8 contain DNA from a non-transfected control, and those labelled 'hit' DNA from the 'hit' clone from which the 'run' clones were derived. The remaining pairs of lanes contain DNA from RUN4 clones.

8.5.3 Conclusions

These results suggest that the RUN4 clones do have an altered methylation status to that of the 'hit' clone within the region of the Hygromycin-*tk* fusion gene. This methylation was not confined to the Hygromycin-*tk* fusion gene as methylated sequences were also detected in the 5' and 3' sequences flanking the integrated 'hit and run' targeting vector. The observation that these 5' and 3' regions were also methylated in non-transfected control cells, indicates that this region of *Cftr* is usually methylated. The methylation of this region of *Cftr* is not unique to ES cells, as a Southern blot conducted by Dr Donald Davidson utilising the same digests and probe combination on DNA from different mouse strains has also shown the pattern of methylated *Hpa II* fragments (figure 8.6). Therefore it appears that the sequences surrounding the target site are normally methylated in mouse cells.

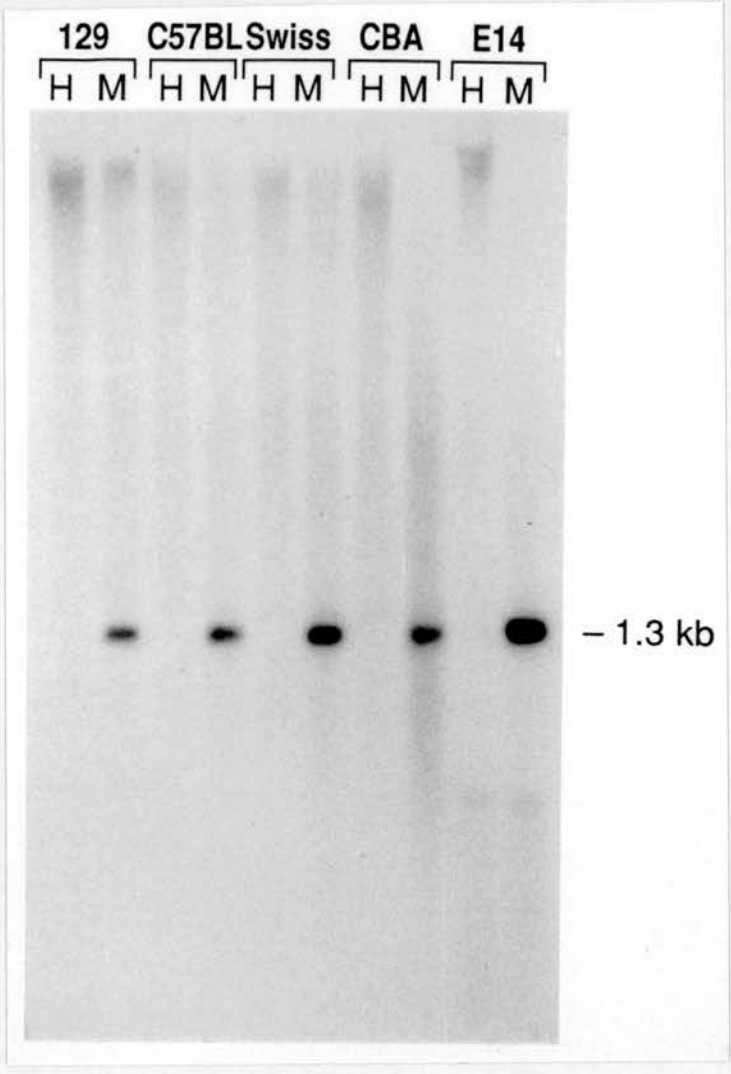


Figure 8.6 Southern blot analysis of the 3' region of the *Cftr* target site in different mouse strains using methylation sensitive enzymes.
This Southern blot was carried out by Dr Donald Davidson using DNA extracted from tail biopsies. The lanes labelled H refer to DNA digested with the methylation sensitive *Hpa II*, and the lanes labelled M to that digested with the methylation insensitive enzyme *Msp I*. The two lanes labelled E14 indicate DNA from an ES cell line, and the remaining lane pairs DNA from different mouse strains.

Although this analysis suggests that the Hygromycin-*tk* fusion gene and flanking regions are methylated, it is by no means conclusive. If time had permitted, further experiments would have been conducted to confirm these observations. In order to determine that the *Sal I* site was intact in those cells whose DNA did not cut at the site present in the targeting vector, the region of DNA containing this site could be amplified by PCR. This amplification would copy the DNA sequence but not the methylation, and therefore digestion of the resultant PCR product with the *Sal I* restriction enzyme should determine whether this site was intact in these cells. A comparison of clones whose selection resistance is indicative of expression or inactivation of the fusion gene by Northern blot analysis should establish the presence or absence of a mRNA transcript from this gene and provide more direct evidence of any deficiencies in gene expression. The absence of bands, or appearance of larger bands in the *Hpa II* digests when compared to the *Msp I* digests might be attributed to incomplete or poor digestion of the DNA by the *Hpa II* restriction enzyme. This possibility could be investigated by reprobing the filters with a probe for a gene which is constitutionally non-methylated in ES cells, e.g. Dihydroxyfolate reductase (Kafri et al. 1992). If the *Hpa II* restriction enzyme was working efficiently and had cut all non-methylated DNA to completion, identical bands should be seen between the *Sal I* and the *Msp I* digested DNA. However despite the absence of this confirmatory data, these results do suggest that the Hygromycin-*tk* fusion gene is methylated in the 'run' clones, in contrast to the 'hit' clone from which they were derived.

Therefore, the difference in the methylation status of the Hygromycin-*tk* fusion gene between the 'hit' clone and the RUN4 clones indicates as suggested previously, that methylation of this region is responsible for the inactivation of the *tk* gene, and also the failure of the restriction enzyme *Sal I* to cut at the site in the *tk* gene. The presence of methylated DNA in the sequences flanking the site of integration of the 'hit and run' targeting vector must also be a contributory factor.

8.6 DISCUSSION

The observation made in the previous chapter that all of the clones surviving Gancyclovir selection still contained the integrated 'hit and run' targeting vector suggested that the *tk* gene of the vector had been inactivated. This proposed inactivation of the *tk* gene produced no obvious band size differences between the Gancyclovir resistant 'run' clones and the 'hit' clone on a Southern blot. This loss of gene activity occurred at a relatively high frequency, and the high numbers of Gancyclovir resistant clones generated could be masking any which had arisen through the desired mechanism of vector excision.

The experiments conducted in this chapter have shown that the RUN4 clones exhibit a variety of selection resistant phenotypes which are associated with a functional or inactive Hygromycin-*tk* fusion gene. These clones appear to be mixed populations with varying levels of cells exhibiting fusion gene activity. The degree to which clone DNA would cut at the *Sal I* restriction site located in the region of the Hygromycin-*tk* gene encoding Gancyclovir sensitivity appeared to correlate with Gancyclovir sensitivity. Resistance to Gancyclovir reflects loss of expression of this gene and was accompanied by a failure of *Sal I* to cut at this site. The mixed population of clones with respect to expression of the Hygromycin- *tk* fusion gene was demonstrated by the ability of the same clone to generate both Gancyclovir resistant cells and Hygromycin resistant cells, but never any cells which were resistant to both selections when imposed simultaneously. This implies that the fusion gene is functional but the activity is being switched on or off.

The observation that clones which had previously been selected on the basis of their resistance to Gancyclovir could exhibit a sensitive phenotype, suggests either that the selection had allowed cells expressing the *tk* gene to survive, or that the mechanism of *tk* gene inactivation which had allowed these clones to survive the original selection was reversible. Various observations confirmed the effectiveness of the Gancyclovir selection, and the presence of cells with differing selection resistance within the same clone population suggests that the latter might apply.

The failure of the restriction enzyme *Sal I* to cut at the site in the region of the Hygromycin-*tk* fusion gene that encodes resistance to Gancyclovir only in those cells which had been selected in Gancyclovir suggested that the mechanism which was responsible for loss of *tk* gene expression was also preventing the enzyme from cutting. As mentioned previously, *Sal I* is a restriction enzyme which is known to be unable to cut methylated CpG sequences. Many studies have correlated repression of gene transcription with methylation of the cytosine of CpG dinucleotides which are commonly clustered at the 5' end of genes, termed 'CpG islands' (Bird 1986; and reviewed in Bird 1992; and Eden and Cedar, 1994). The exact mechanism by which methylation represses gene expression is not known but it is thought that methylation might suppress gene transcription by direct inhibition of transcription factor binding (Becker *et al.* 1987), or indirectly through Methyl CpG-binding proteins (MeCPs) which specifically bind to multiple symmetrically methylated CpGs (Boyes and Bird, 1991). Increased DNA methylation and decreased gene expression have also been found to correlate with the formation of a 'closed' chromatin conformation which may be inaccessible to the transcriptional machinery (Antequera *et al.* 1990).

Methylation of DNA is an epigenetic mechanism of gene alteration which does not involve actual mutation of the DNA and is consequently reversible. It is therefore possible that methylation-induced inactivation of the Hygromycin-*tk* fusion gene is generating Gancyclovir resistant clones during selection for the 'run' event. Some or all of the cells in this clone population could subsequently 'lose' this methylation when the selection pressure for this phenotype is removed, which would in turn remove the suppression of gene expression leading to a Gancyclovir sensitive phenotype. Cooper *et al.* (1992) have described a similar 'switching' phenotype in a mouse teratocarcinoma cell line, where clones lost and reacquired expression of the *aprt* (adenine phosphoribosyltransferase) gene involved in the purine salvage pathway. Analysis of these clones found that loss of *aprt* activity correlated with reduction of *aprt* mRNA, resistance to nuclease digestion which is characteristic of chromatin formation, and increased CpG island methylation, all of which are

indicative of transcriptional suppression by methylation. The *hsvtk* gene is also susceptible to methylation-induced gene inactivation. Graessmann *et al.* (1994) have reported that methylation of promoter sequences of the *hsvtk* gene and even of non-promoter CpG sequences located downstream in the coding sequences can suppress its transcription. Therefore loss of Thymidine kinase gene activity through methylation-induced suppression of the Hygromycin-*tk* fusion gene expression could be an explanation for the variable Gancyclovir resistance of the 'run' clones.

A subsequent investigation into the methylation status of the RUN4 clones showed that there were indeed differences between the 'hit' clone and the 'run' clones which had been selected in Gancyclovir. The Hygromycin-*tk* fusion gene of the 'run' clones did appear to be methylated whereas that of the 'hit' clone did not. Therefore it appears that Gancyclovir selection had selected cells in which the selection gene had become methylated resulting in loss of *tk* gene expression.

The high frequency at which the Hygromycin-*tk* gene became methylated might be explained by the observation that the genomic sequences flanking the target site are usually methylated in mouse cells. The phenomenon of loss of gene expression when placed adjacent to methylated sequences has been well documented and Kass *et al.* (1993) have shown that this is probably a result of spread of chromatin formation from the methylated region into an adjacent non-methylated region, thereby inactivating any gene which is present. It has been proposed that methylation is not a prerequisite for chromatin formation, and that methylation of DNA alone is not sufficient to block *hsvtk* gene transcription (Buschhausen *et al.* 1987). The methylation of DNA sequences which have been converted into an inactive chromatin form has been suggested to occur during replication (Kass *et al.* 1993). Therefore it is possible that the methylated sequences flanking the integrated 'hit and run' targeting vector were in the inactive chromatin form which then spread into the Hygromycin-*tk* fusion gene in some cells, thereby suppressing its transcription. These cells would then become resistant to Gancyclovir and would be selected for during the 'run' step. Replication of the fusion gene in this inactive chromatin

formation in these Gancyclovir resistant clones might then result in its subsequent methylation which was detected in this investigation by use of methylation-sensitive enzymes. This theory could be confirmed further by screening the region of the *Cftr* target site of these clones for the presence of sequences in the inactive chromatin form which are characterised by a resistance to *Msp I* nuclease digestion (Keshet *et al.* 1986). The chromatin structure of the promoter region of CFTR has been studied (Koh *et al.* 1993), and its presence was found to correlate with expression levels in cell lines *in vitro*. However, it is unlikely that the methylation and proposed chromatin structure of sequences flanking exon 10 identified in this study exert an effect upon expression of CFTR.

If the failure of these experiments to obtain clones in which the vector has excised is due to the high background of methylation-induced loss of *tk* gene expression, this does not explain why clones are obtained which appear to have excised the ΔF508 vector in similar experiments conducted in parallel in our lab. Since the ΔF508 'hit and run' targeting vector utilises the same negative selection to select for loss of the *tk* gene expression, and is integrated into the same region of *Cftr*, it would be expected to also be susceptible to the same level of methylation-induced loss of *tk* gene expression. The Gancyclovir resistant clones which appeared to have excised the ΔF508 'hit and run' targeting vector had always reverted to wildtype, and did not retain the ΔF508 mutation. Although the ratio of duplicated sequences does favour loss of the mutation, retention of the ΔF508 mutation would have been expected in 20% of the clones. It is possible that the wildtype clones obtained with these ΔF508 'hit' clones represent wildtype cells which were already present before Gancyclovir selection was imposed. The presence of such 'wildtype contamination' in the 'hit' clone would not be apparent until a selection was applied which selected for cells which did not contain the targeting vector. Due to the expected precision of the excision process, it would not be possible to distinguish between wildtype cells and cells which had previously been targeted but which had subsequently excised the vector. The main difference between the ΔF508 'hit and run' targeting vector and the ΔI506/7 vector is the positive selection gene. The ΔF508 vector utilises a Neomycin

resistance gene to select for vector expression, whereas the ΔI506/7 utilises Hygromycin. The absence of any wildtype cells from the Gancyclovir selection of ΔI506/7 'hit' clones might reflect the difference in stringency of the positive selections, with Hygromycin being more effective at eliminating any non-expressing cells. The targeted ΔF508 'hit' clone which had given rise to the wildtype 'run' clones was subsequently recloned by plating some cells out to generate clones from single cells. Analysis of the resulting clones found that only 2 out of 7 clones contained the targeted 'hit and run' vector, indicating that the ΔF508 'hit' clone had a high level of contamination with wildtype cells. This would account for the high number of Gancyclovir resistant clones obtained from this 'hit' which had appeared to have excised the 'hit and run' targeting vector. Therefore, the similar ΔF508 'hit and run' targeting vector could be susceptible to the same mechanism of methylation-induced gene inactivation experienced by the ΔI506/7 'hit and run' targeting vector. However this effect had been masked the high level of wildtype cells in the 'hit' clone, which had given the impression that Gancyclovir selection had selected cells in which the vector had excised. Subsequent Gancyclovir selection on the recloned ΔF508 'hit' clone generated 'run' clones in which the 'hit and run' targeting vector was present in all Gancyclovir resistant clones suggesting that this vector was also subject to the same loss of *tk* gene expression as the ΔI506/7 targeting vector.

Therefore it appears that the inactive chromatin of methylated sequences flanking the integrated 'hit and run' targeting vector had spread into the Hygromycin-*tk* fusion gene at a high frequency during the 'run' stage. This inactivated its expression, thereby rendering cells resistant to the Gancyclovir selection and enabling them to survive the selection which had been imposed to select for cells which had lost *tk* gene expression through vector excision. This does not exclude the possibility that some cells are excising the targeting vector to become Gancyclovir resistant, as the high background of methylated Gancyclovir resistant clones would obscure such an event if it were occurring at a lower frequency.

CHAPTER 9

MODIFICATION OF THE 'HIT AND RUN' TARGETING

9.1 INTRODUCTION

The experiments described so far have been unsuccessful in selecting clones which have survived Gancyclovir selection through excision of the integrated 'hit and run' targeting vector. The 'hit and run' technique was modified in light of the observations made in previous experiments to try and overcome the problems experienced to date.

9.2 RUN 7 'RUN' EXPERIMENT

The failure of Gancyclovir selection to select any cells which had lost *tk* gene expression through excision of the 'hit and run' targeting vector was thought to be due to the high background of Gancyclovir resistant clones generated by a more frequent mechanism of gene inactivation, obscuring the less frequent vector excision event. An investigation into this phenomenon had shown that methylation-induced suppression of *tk* gene expression was responsible for generating high numbers of Gancyclovir resistant clones.

To try and cut down this high level of 'false positives', Gancyclovir resistant clones obtained from selection for a 'run' event were screened for resistance to Hygromycin. Clones which have become resistant to Gancyclovir through methylation inactivation of the *tk* gene are likely to be mixed populations of expressing and non-expressing cells, and so at least some cells from such a clone would grow in Hygromycin selection. Clones which are resistant to Gancyclovir through vector excision would not be able to reverse their phenotype and so would remain completely sensitive to Hygromycin selection. This method of testing the Hygromycin sensitivity of Gancyclovir resistant clones should allow the identification and elimination of those still containing the targeting vector at an early stage, without the need to grow them up for analysis. The removal of a requirement to grow up cells to obtain DNA for Southern blotting in order to identify 'false positives' will allow more clones to be picked and analysed, and therefore increases the chances of detecting a true 'run' event.

9.2.1 Method

A passage 30 culture of the HR5 'hit' clone 1.10 was grown up on Hygromycin selection for 7 days and then used to set up a 'run' experiment. Cells were seeded at a density of 1 X 10^5 cells per 100 mm plate onto twenty five plates containing a PEF feeder layer, and seventy five plates in the absence of a feeder layer. Gancyclovir selection was imposed onto half of the plates two days after plating, and the other half three days after plating. The medium was replaced every three days until clones were visible.

Once visible, clones were picked and screened by the following procedure (illustrated in figure 9.1). The Gancyclovir resistant clones were picked, trypsinised, and seeded between two wells of a ninety six well plate, one of which contained a PEF feeder layer. Hygromycin selection was imposed upon the cells seeded in the absence of a feeder layer, and the other well was maintained without selection. After four days, the cells in Hygromycin selection were scored for Hygromycin resistance. Clones in the wells which were not on Hygromycin selection were taken further only if all of the cells in the corresponding well died in Hygromycin selection. Clones which did grow in Hygromycin were discarded, as these were likely to represent cells which still contain the 'hit and run' targeting vector. Clones which appeared to be Hygromycin sensitive, were split into two more wells of a ninety six well plate when confluent, and Hygromycin selection was imposed upon one of these wells for a second time. This second Hygromycin step served to eliminate those clones which had either appeared to be Hygromycin sensitive due to a failure of cells to grow for reasons such as too low a cell density, or which had not yet reversed the DNA methylation. Clones which still appeared to be completely sensitive to Hygromycin selection were then grown up to generate cells for frozen stocks and DNA analysis.

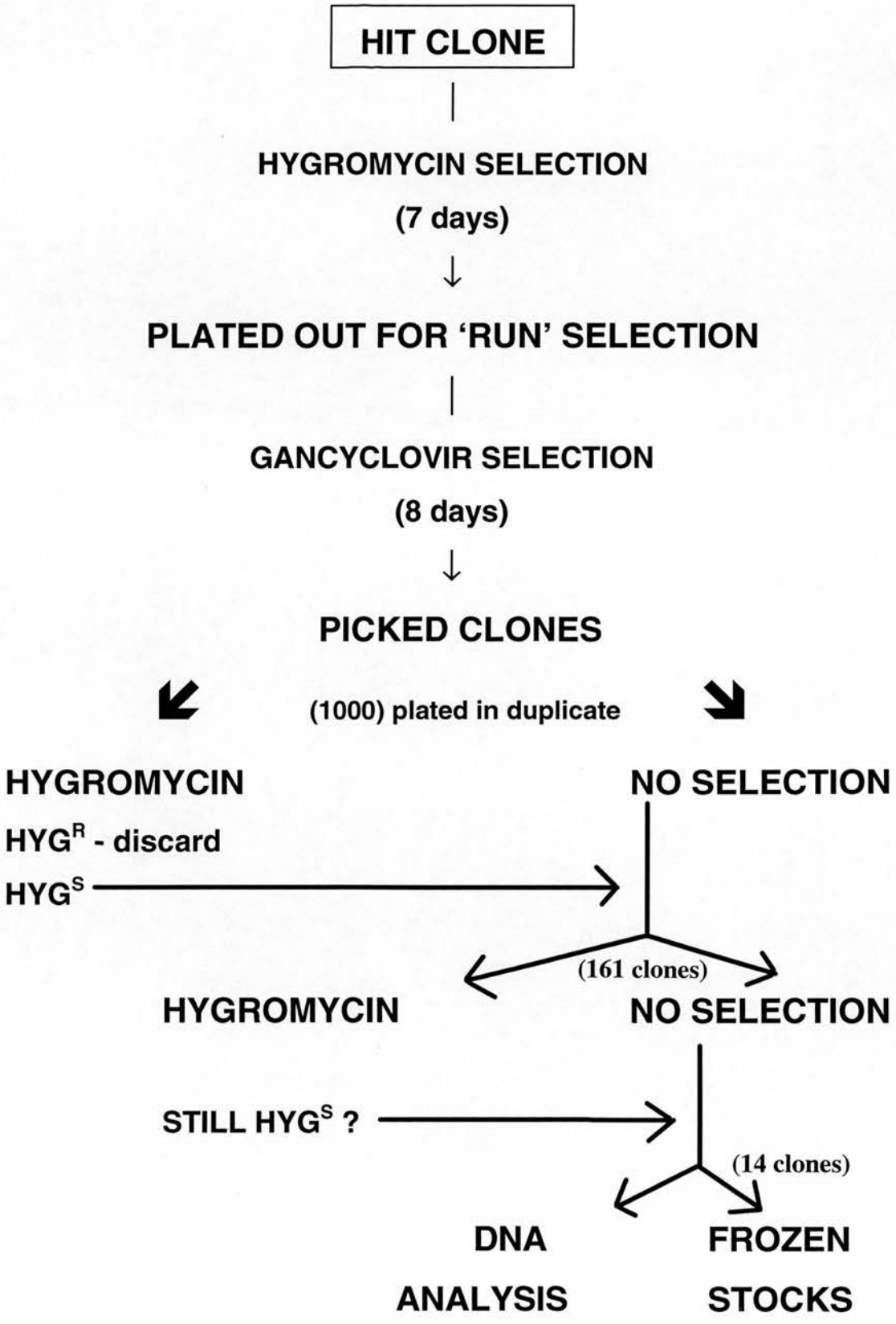


Figure 9.1 Flow chart illustration of the modified 'run' selection procedure.

9.2.2 Results

9.2.2.1 Clone numbers

A large number of Gancyclovir resistant clones were generated in this experiment from the plates in which cells were seeded onto PEF feeder layers. Plates in which cells were plated in the absence of a feeder layer generated a very small number of clones, considering the number of plates that were seeded (12 clones from 75 plates). There did not appear to be any difference between the number of clones obtained from the plates on which Gancyclovir selection was added two days after plating than the plates on which it was imposed after three days. The number of Gancyclovir clones which were picked are detailed in table 9.1.

One thousand Gancyclovir resistant clones in total were picked and screened for Hygromycin resistance. The majority of clones were found to be Hygromycin resistant and were discarded leaving 161 clones which appeared to be sensitive to Hygromycin to be screened a second time. This number of clones was further reduced by the second round of screening to just 14 which were consistently sensitive to Hygromycin. These clones were grown up and analysed by Southern blot.

Table 9.1 The number of Gancyclovir resistant clones picked in experiment RUN7.

	FEEDER LAYER ?	TIME OF GANC ADDITION (days)	No. CLONES PICKED
	+	2	5
	+	2	4
	+	2	6
	+	2	9
	+	2	2
	+	2	2
	+	2	40
	+	2	27
	+	2	11
	+	2	7
	+	2	67
	+	2	90
	+	2	36
	+	2	72
	+	2	72
	+	2	96
TOTAL	—	—	**546**
	+	3	58
	+	3	48
	+	3	49
	+	3	36
	+	3	39
	+	3	96
	+	3	98
	+	3	7
	+	3	11
TOTAL	—	—	**442**
	-	2	2
	-	2	1
	-	3	3
	-	3	2
	-	3	4
TOTAL	—	—	**12**
GRAND TOTAL	—	—	**1000**

9.2.2.2 Screening RUN7 clones for vector excision.

The Hygromycin sensitive, Gancyclovir resistant clones were analysed by probing *Xba I-Sal I* digests with the 0.6EH probe which had been used previously to screen for 'hit' clones in section 6.2.2.2 (figure 9.2). The presence of the integrated 'hit and run' targeting vector in the target site is indicated by the appearance of an 8 kb band above the 6 kb band of the wildtype allele. Methylation of the *Sal I* restriction site present in the 'hit and run' vector inhibits *Sal I* from cutting at this site, and therefore the 16.5 kb band which is generated by insertion of the targeting vector into the 6 kb *Xba I* fragment is not reduced to 8 kb.

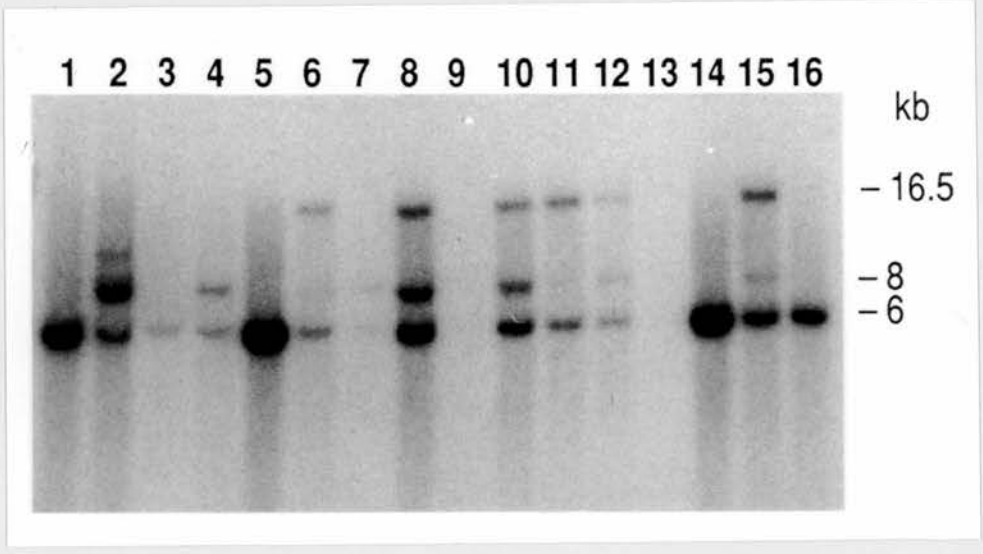


Figure 9.2 Screening RUN7 clones by Southern blot for vector excision.
DNA was digested with *Xba I* and *Sal I* and probed with 0.6EH probe. Lane 1 is the non-transfected control and displays the single 6 kb wildtype band. Lane 2 is the 'hit' clone which has not been subjected to Gancyclovir selection and shows the 8 kb band created by the integrated targeting vector as well as the 6 kb band of the wildtype allele. Lanes 3 to 16 are the RUN7 clones. Lanes 3, 5, 14, and 16 contain RUN7 clones displaying a single wildtype band indicative of vector excision.

When screened by this method, eight clones gave one or both of the 8 kb or 16 kb bands indicative of the presence of the 'hit and run' targeting vector at the target site. However, in contrast to all previous 'run' experiments, four clones did not have any vector bands, and displayed just a single wildtype band. These four clones therefore appear to be 'run' clones which have excised the 'hit and run' targeting vector. Three of these clones originated upon plates which had been seeded with a PEF feeder layer, and one from a plate which did not contain a PEF feeder layer. Three of the four clones occurred upon different plates, indicating that they are not clonal and represent at least three independent recombination events.

9.2.2.3 Screening RUN7 clones for retention of ΔI506/7 mutation.

The RUN7 clones were screened to determine whether vector excision had resulted in loss of the mutation and reversion to a wildtype genotype, or whether the mutation had been retained as the only alteration to *Cftr*.

Clone DNA was analysed by the same combination of digests and probe used to identify the ΔI506//7 mutation in the targeted 'hit' clones (section 6.2.2.4), i.e. digested with *Ssp I* and *Eco RI*, blotted and probed with the 0.8HE probe. The presence of the ΔI506/7 mutation in the mutant exon introduces a unique *Ssp I* site which generates a novel 0.8 kb band in addition to the 2.2 kb band generated by the wildtype and non-mutant exon.

Analysis of the RUN7 clones showed that none of the clones which had excised the vector had the 0.8 kb band indicating that they had not retained the ΔI506/7 mutation, but had instead reverted to a wildtype status (figure 9.3).

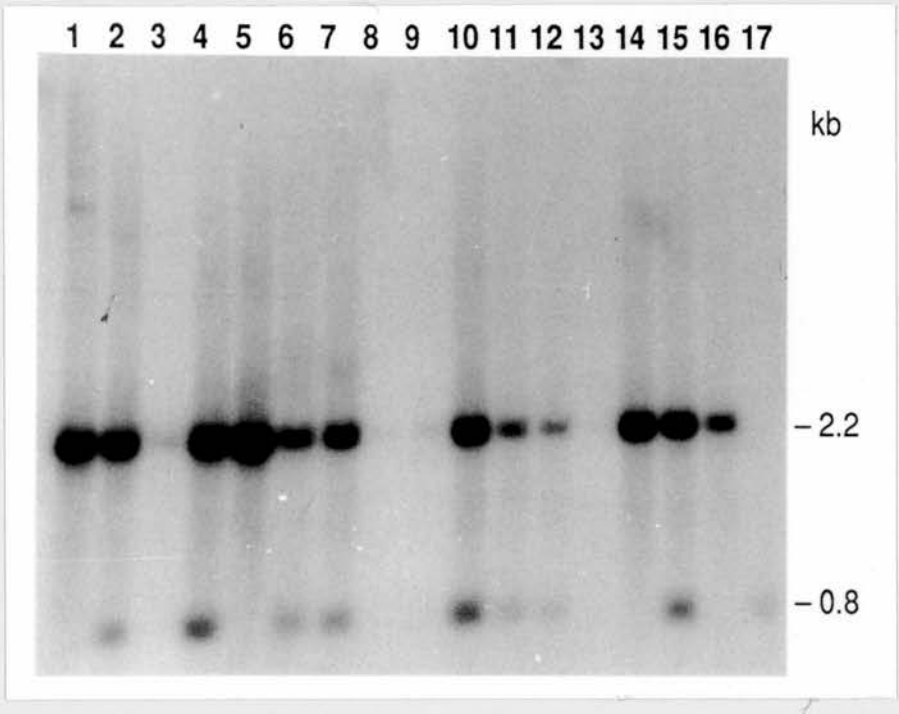


Figure 9.3 Screening RUN7 clones by Southern blot analysis for retention of the ΔI506/7 mutation.

A Southern blot of *Eco RI* and *Ssp I* digested DNA was probed with the 0.8HE probe. Lane 1 is the non-transfected control displaying the 2.2 kb wildtype band. Lane 2 is the 'hit' which displays the 2.2 kb band generated by both the wildtype allele and the non-mutant exon, along with the 0.8 kb band generated by the *Ssp I* site of the mutation. (For a restriction map of these sites see figure 6.1). Lane 17 contains vector DNA only, demonstrating the 0.8 kb band generated by the mutation carried by the vector sequences. Lanes 3 to 16 are the RUN7 clones, with lanes 3, 5, 14, and 16 containing DNA from the clones which had excised the targeting vector.

9.2.3 Conclusions

This experiment has, for the first time generated clones which appeared to have become Gancyclovir resistant through excision of the 'hit and run' targeting vector, rather than through methylation-induced inactivation of the *tk* gene.

The results from previous experiments suggested that methylation-induced inactivation of the *tk* gene is a common event in the targeted 'hit' clones, and that all of the Gancyclovir resistant clones obtained to date had survived Gancyclovir selection through this mechanism. It was proposed that the high numbers of these methylated clones could be masking any which had become Gancyclovir resistant through the desired mechanism of vector excision if this were occurring at a much lower frequency. This experiment described an attempt to identify and eliminate these methylated clones at an early stage, allowing a greater number of Gancyclovir resistant clones to be screened, so that only those unlikely to be methylated would be taken through to the analysis stage.

This modified experiment was successful in identifying four clones in which the vector appeared to have excised. The majority of the remaining clones displayed the 16.5 kb band indicative of methylation of the *Sal I* site and probably represent clones which are still methylated to such an extent that they remain completely sensitive to Hygromycin. One clone did not display this 16.5 kb methylated band and perhaps represents a clone in which the Hygromycin-*tk* has been inactivated by some other mechanism, such as a point mutation in the promoter region.

Of the four 'run' clones obtained, three had been selected upon PEF feeder layers, and one in the absence of a feeder layer. As many more Gancyclovir resistant clones were generated from the cells seeded onto PEF feeder layer plates than were analysed, an absolute frequency of vector excision cannot be calculated for these clones. However, all of the clones arising on plates seeded in the absence of feeder layers were analysed, and therefore a frequency of the 'run' event under these conditions can be determined. These plates were seeded with a total of 7.5×10^6

cells, and gave rise to twelve Gancyclovir resistant clones, of which one had appeared to have excised the 'hit and run' targeting vector. This gives a frequency of excision of the targeting vector in 1.3 X 10^{-7} cells plated. However, Gancyclovir selection was imposed upon this plate 2 days after removal of Hygromycin selection, and therefore this figure must be corrected for the increase in cell numbers during this intervening period. The mean doubling time for ES cells growing exponentially is approximately once every eighteen hours. In the 2 days between the different selections, the cells should have doubled approximately 2.6 times. Therefore it can be estimated that the cell numbers at the time of Gancyclovir selection were likely to be 4.8 X 10^7 and therefore the frequency of vector excision would be 1 in 4.8 X 10^7 i.e. 2 X 10^{-8} cells. Although only a small proportion of the Gancyclovir resistant clones arising on PEF feeder layers were analysed, a 'minimum' frequency of vector excision can be calculated as we know that at least three 'true run' clones were generated. The total number of cells seeded under these conditions was 2.5 X 10^6, of which 14 plates had Gancyclovir added after 2 days, and 11 plates had it added after 3 days. Therefore after correcting for an overall increase in cell numbers to 2.9 X 10^7, it can be estimated that the frequency of vector excision under these conditions must occur in at least 3 in 2.9 X 10^7 cells, i.e. 1.0 X 10^{-7} cells plated, and is likely to be higher. This frequency is at least tenfold higher than for cells selected in the absence of a PEF feeder layer. However far fewer Gancyclovir resistant clones were generated on the plates which did not contain PEF feeder layers and therefore despite the lower frequency of vector excision, it might be preferable to select cells in the absence of PEF feeder layers in future experiments, as this lower 'false positive' background should facilitate identification of true 'run' clones. The higher numbers of Gancyclovir resistant clones, and the increased frequency of vector excision in cells selected upon PEF feeder layers could simply be a reflection of the increased plating efficiency under these conditions, or a reduced susceptibility to the 'bystander effect' conferred by the more compact growth of ES cells upon PEF feeder layers. This frequency of vector excision is comparable to that of 8 X 10^{-7} per cells plated reported by Valancius and Smithies (1991), and 4.3 X 10^{-6} reported by Hasty *et al.* (1991), when targeting the *hprt* locus. A higher frequency of 3.8 X 10^{-3} was reported

by Hasty *et al.* (1991) when targeting the *Hox 2.6* gene, suggesting that the frequency of this recombination event may vary between loci.

Further analysis showed that the ΔI506/7 mutation was absent from the clones which had appeared to have lost the targeting vector, indicating that the mutant exon had excised along with the vector reverting the clones back to a wildtype status. Due to the expected precision of the excision event, it is not possible to distinguish between clones which have excised the 'hit and run' targeting vector and wild type cells which have not been targeted. Therefore the possibility cannot be excluded that these wildtype clones were not derived from targeted 'hit' cells but are the result of a low level contamination of the 'hit' clone with wildtype cells. This however is unlikely, as any wildtype cells should have been eliminated from the 'hit' clone population during the Hygromycin selection which preceded the Gancyclovir selection. The frequency of any wildtype cells in the 'hit' clone would be expected to be much greater if the hit clone had not be subjected to Hygromycin selection for some period of time. Therefore, the RUN2 and RUN3 experiments which utilised cells from a 'hit' clone which had been grown for a long period without Hygromycin selection were likely to have a greater number of these wildtype cells in the population. Although ten fold fewer Gancyclovir resistant clones were analysed in these experiments, any wildtype cells which were present in the 'hit' clone population would have been allowed to increase in numbers for some time, and would have been expected to reach a detectable level. The complete absence of any clones with a wildtype genotype in previous experiments suggests that the 'hit' clones were not contaminated with a low level of wildtype cells. The only irrefutable evidence that these clones had excised the targeting vector however, would be the retention of the ΔI506/7 mutation.

The observation that all four 'run' clones appeared to have lost the mutation and reverted back to a wildtype genotype was not entirely unexpected. The distribution and size of the duplicated genomic regions is such that the region through which recombination would result in retention of the ΔI506/7 mutation is smaller than the

region through which recombination results in loss of the mutation, and therefore mutation loss is favoured (figure 9.4). If recombination occurred at the same frequency along the entire length of these DNA sequences, the size difference between the two genomic regions of 0.7 kb compared to 3.5 kb would be expected to result in retention of the ΔI506/7 mutation in 1 in 6 (17%) recombination events. Therefore the observation that all four clones had lost the mutation is consistent with the frequency of mutation loss predicted by the genomic structure.

This experiment was therefore successful in identifying the majority of clones which were resistant to Gancyclovir through methylation-induced inactivation of the fusion gene, without the need for Southern blot analysis. As predicted, the elimination of such clones at an early stage enabled a greater number of clones to be screened, thus allowing the identification of the very small number of clones which appeared to Gancyclovir resistant through vector excision. Unfortunately none of the clones which had excised the targeting vector retained the mutation, although this was expected for such a small number of clones as the genomic structure of the targeted allele favours mutation loss.

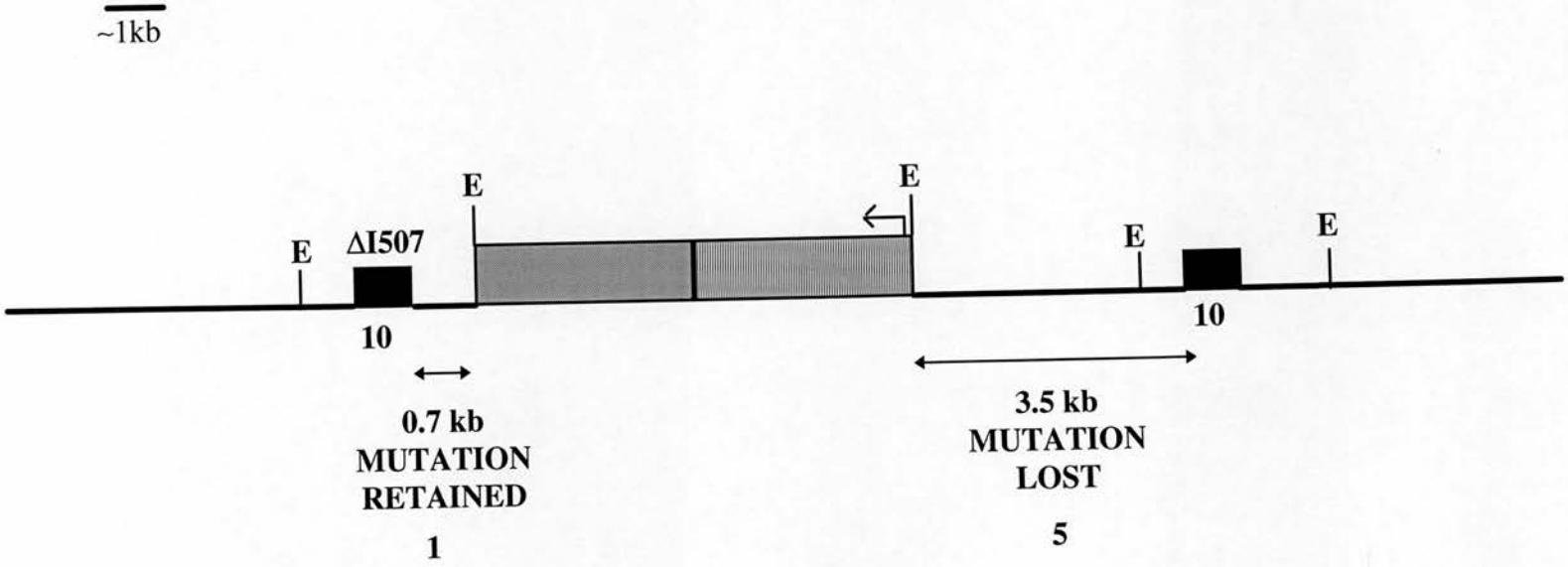


Figure 9.4 Duplicated regions available for intrachromosomal recombination during a 'run' event.
The plasmid sequences are indicated by the diagonally hatched area, and the Hygromycin-*tk* fusion gene by the vertically lined region. The direction of transcription is indicated by the bent arrow. Exon 10 is represented as a dark rectangle, with the presence of the ΔI506/7 mutation indicated above the mutant exon. The extent of the duplicated genomic regions is indicated by the double-headed arrows, with the outcome of a recombination event within each region indicated below. The restriction sites for *Eco RI* are indicated by E.

9.3 TARGETING *Cftr* WITH A MODIFIED 'HIT AND RUN' TARGETING VECTOR

The 'hit and run' targeting vector used in all previous experiments favours loss of the mutation during the 'run' step due to the greater length of the region through which recombination results in loss of the mutant compared to that which results in excision of the endogenous (non-mutant) exon (figure 9.4). A new vector was constructed by Dr Paul Dickinson which has an increased region available for the recombination event which would result in retention of the mutation. This was used to target *Cftr* to generate 'hit' clones which could then be used to obtain 'run' clones which should have a greater chance of retaining the mutation upon vector excision.

9.3.1 Method

9.3.1.1 Vector construction

The modified 'hit and run' targeting vector incorporated an additional 1.8 kb of homologous genomic sequences which were inserted as an *Eco RI* fragment into the *Eco RI* site just downstream of exon 10 (figure 9.5). This increases the region of homology through which recombination will result in retention of the ΔI506/7 mutation from 0.7 kb to 2.5 kb, and although mutation loss is still favoured, the mutation should be retained in 42% of 'true run' clones (assuming an equal frequency of homologous recombination along the length of targeting DNA).

The modified 'hit and run' targeting vector does not utilise the Hygromycin-*tk* fusion gene for selection resistance, but instead utilises the separate Neomycin and Thymidine kinase genes expressed from separate *tk* promoters which had been used previously in the ΔF508 'hit and run' targeting vector.

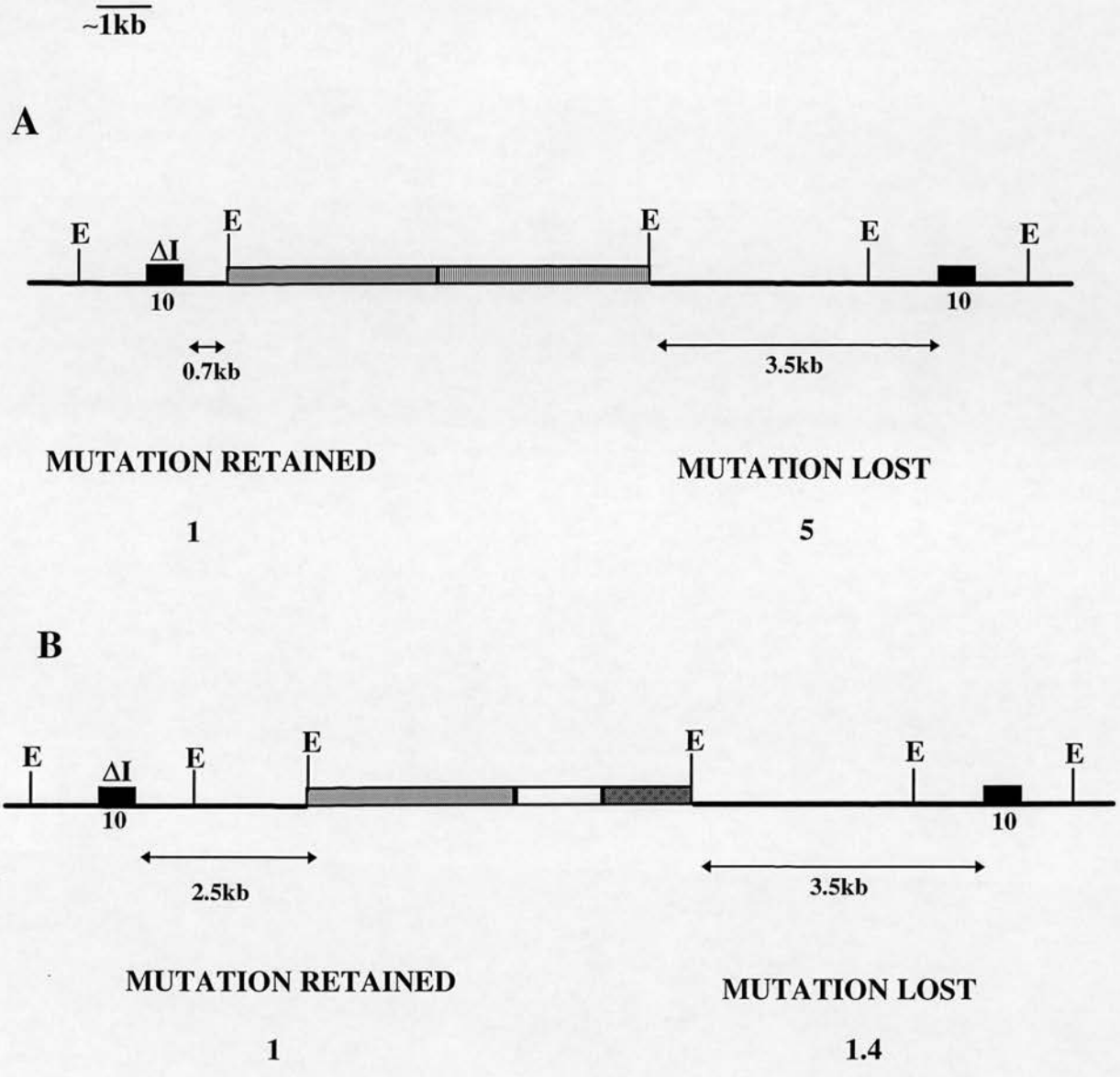


Figure 9.5 Distribution of genomic sequences of 'hit and run' targeting vectors. (A) Structure of 'hit and run' targeting pHTHRΔI507 vector used in previous experiments. (B) Structure of 'hit and run' targeting vector pHRΔI which incorporates an additional 1.8 kb of homology to the target site. The genomic sequences are indicated by the thick lines and exon 10 by the dark rectangle. The presence of the ΔI506/7 mutation is indicated by ΔI. Plasmid sequences are denoted by a diagonally hatched region, the Hygromycin-*tk* fusion gene by a vertical lined area, the neomycin gene by an area of trellis hatching, and the separate *tk* gene by the open rectangle. The genomic regions available for recombination during a vector excision event are indicated by the arrowed lines, with the outcome of each event indicated beneath. The restriction sites for *EcoRI* are denoted by E.

9.3.1.2 Targeting experiment HR6.

A passage 10 culture of CGR8 ES cells was grown up on PEF feeder layer, and 1 X 10^7 cells transfected by electroporation with 100 μg of *Hpa I* linearised targeting vector pHRΔI. The cells were plated into ten plates in the absence of a PEF feeder layer at a density of 1 X 10^6 cells per plate, and G418 selection was applied at a concentration of 200 μg/ml twenty four hours after transfection. The plates were fed every three days, and once visible, clones were picked and grown up for analysis.

9.3.2 Results

9.3.2.1 Clone numbers.

The numbers of clones generated by this transfection were much lower than in previous experiments using the 'hit and run' targeting vector pHTHRΔI506/7. Seventy six G418 resistant clones were obtained in total, giving a transfection efficiency of 7.6 X 10^{-6}. From the seventy six clones obtained, 32 were taken through to Southern blot analysis.

9.3.2.2 Southern blot analysis for targeted clones.

The G418 resistant clones were analysed by Southern blot to detect those clones in which the targeting vector had integrated through homologous recombination into the target site. Clone DNA was digested by the enzyme *Nco I*, blotted and probed with the 1.3XH probe. This detects a 20 kb fragment of the wildtype allele and a 18 kb fragment generated by integration of the targeting vector into the target site (figure 9.6).

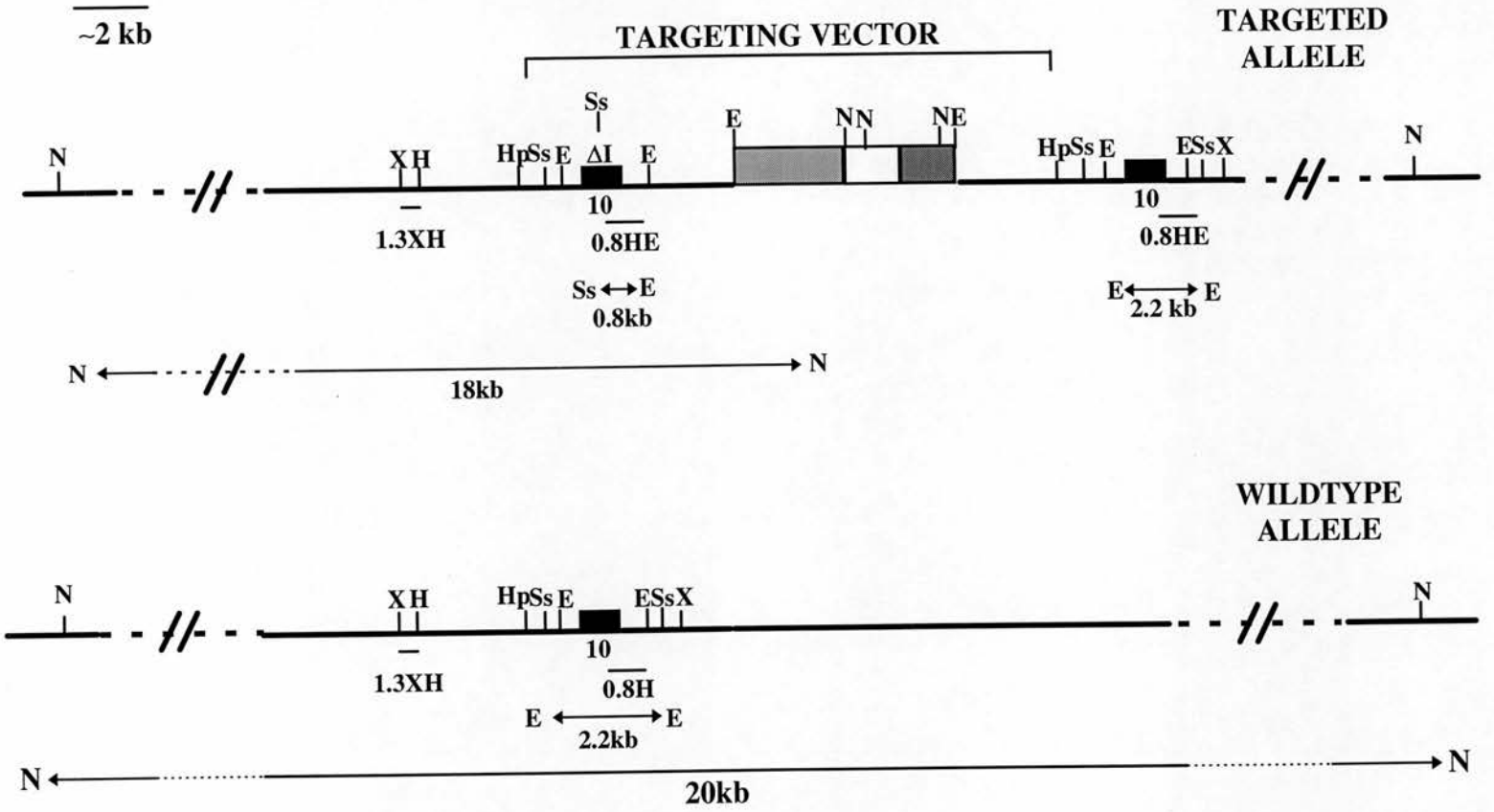


Figure 9.6 Restriction map of *Cftr* alleles of clones targeted with pHRΔI.
Plasmid sequences are indicated by the diagonal hatching, the *tk* gene by the open rectangle, and the neomycin gene by the trellis hatching. Genomic sequences are denoted by the thick line, exon 10 by the closed rectangle, and the position of the ΔI506/7 mutation by ΔI. The dotted lines indicate a region of condensed scale. The sites of hybridisation of each probe are indicated, and the bands produced are represented by the double headed arrows. The restriction sites are as follows: *EcoRI* (E), *Hpa I* (Hp), *Hind III* (H), *Nco I* (N), *Ssp I* (Ss), and *Xba I* (X).

Analysis of the HR6 clones detected 4 clones which displayed both 20 and 18 kb bands indicating that they were homologous recombinants (figure 9.7). The identification of 4 correctly targeted clones out of 32 clones analysed gives a targeting frequency of 1 in 7 (14%). All four clones originated upon different plates, indicating that they arose from four independent homologous recombination events. The wildtype and mutant bands were present at equal intensity, indicating the correct dosage of a single copy of both the mutant and wildtype alleles.

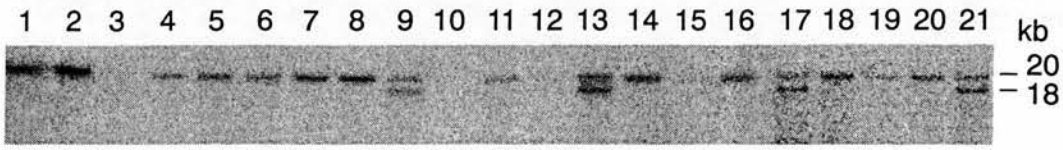


Figure 9.7 Screening experiment HR6 for targeted 'hit' clones.
Southern blot analysis of HR6 clone DNA digested with *Nco I* and probed with the 1.3XH probe. Lane 1 is the non-transfected control displaying the 20 kb band of the wildtype allele. Lanes 2 to 21 are HR6 clones, with targeted 'hit' clones indicated in lanes 9, 13, 17, and 21 by the presence of the additional 18 kb band created by integration of the targeting vector into the target site.

9.3.2.3 Screening the 'hit' clones for retention of the ΔI506/7 mutation.

The correctly targeted clones were analysed for the presence of the mutation by Southern blot utilising the same combination of digests and probe used to screen the HR3 and HR5 'hit clones (sections 6.2.2.4 and 6.3.2.4), i.e. *Eco RI* and *Ssp I* digests probed with the 0.8HE probe. This demonstrated that one of the 'hit' clones had retained the ΔI506/7 mutation (figure 9.8) and therefore was suitable for use in the 'run' stage.

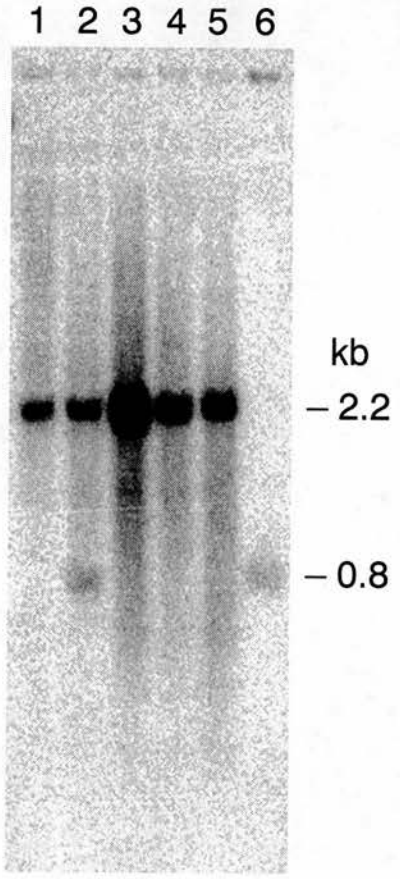


Figure 9.8 Screening HR6 'hit' clones for the presence of the ΔI506/7 mutation. Southern blot analysis of HR6 'hit' clones digested with *Ssp I-EcoRI* and probed with the 0.8HE probe. Lane 1 is the non-transfected control displaying the 2.2 kb band of both the wildtype and non-mutant exon 10. Lane 6 contains vector DNA alone and indicates the 0.8 kb band generated by the mutant exon. Lanes 2 to 5 contain HR6 'hit' clones, with the only clone retaining the mutation present in lane 2.

9.3.3 Conclusions.

This experiment was successful in targeting *Cftr* with a modified version of the 'hit and run' targeting vector used in previous experiments. Despite achieving a lower transfection efficiency, the targeting frequency obtained with this modified vector was the same as that obtained for the targeting experiment HR5. As discussed previously (section 6.4), an increase in the length of the homology to the target site carried by the targeting vector has been shown to increase the targeting frequency at a locus. Therefore this modified 'hit and run' targeting vector which incorporates an additional 1.8 kb of homologous sequence might have been expected to target *Cftr* at a higher frequency than at the same frequency of 14% which had been obtained with the previous vector.

One of these targeted clones had retained the mutation and is suitable for use in the next stage, the selection for a 'run' event. These clones should still be vulnerable to methylation-induced inactivation of the *tk* gene and therefore selection and analysis of 'run' clones should be carried out using the modified procedure described in section 9.2 to eliminate the majority of methylated clones. The increased length of homology carried by these targeted clones should increase the frequency at which the mutation is retained in a vector excision event, and perhaps allow clones to be obtained which have retained the ΔI506/7 mutation as the only alteration to *Cftr*.

9.4 DISCUSSION

This chapter describes two attempts to overcome the problems encountered to date with the 'run' step of the 'hit and run' targeting strategy.

The first approach was to modify the procedure of the 'run' step to identify and eliminate those Gancyclovir resistant clones which had arisen through methylation-induced inactivation of the *tk* gene at an early stage, without the need for Southern blot analysis. By taking advantage of the apparent reversible nature of the fusion gene inactivation, the procedure of screening clones after picking was successful in identifying the majority of methylated clones. As predicted, a reduction in this 'false

positive' Gancyclovir background allowed clones which had excised the targeting vector to be identified.

The frequency at which the vector excised was determined as 2 X 10^{-8} for cells selected in the absence of a PEF feeder layer, and at least 1 X 10^{-7} in cells selected in the presence of a PEF feeder layer. If this frequency of vector excision had been consistent throughout all of the previous 'run' experiments, the numbers of clones analysed would not have been sufficient to have identified a recombination event. Experiment RUN2 utilised 1.2 X 10^{6} cells seeded onto PEFS for the 'run' step and therefore might have generated a single vector excision event. However, less than 5% of the resulting Gancyclovir resistant clones were analysed, and consequently a true 'run' clone was not detected. All other 'run' experiments used too few cells, and did not analyse a large enough proportion of the resultant clones to identify a vector excision event which was occurring at these frequencies. Therefore as predicted, the frequency at which vector excision was occurring in these 'run' experiments was very low, and was obscured by the large number of Gancyclovir resistant clones generated by the higher frequency of methylation-induced inactivation of the Hygromycin-*tk* fusion gene.

The larger number of 'run' clones analysed in this experiment identified four clones which had excised the targeting vector. Unfortunately all of the clones which had excised the targeting vector had reverted back to a wildtype genotype and had lost the mutation. The distribution of the duplicated genomic sequences created by insertion of the 'hit and run' targeting vector into *Cftr* is such that recombination resulting in loss of the mutation is favoured, and retention of the mutation might be expected to occur in approximately 1 in 6 recombination events. Therefore a greater number of clones which had excised the targeting vector would be required to obtain one which had retained the ΔI506/7 mutation. Based on the results of this experiment, this might best be achieved by selecting at least 6 X 10^{9} cells in the absence of a PEF feeder layer, and analysing all of the resulting Gancyclovir resistant clones.

The second approach overcoming the problems encountered too date was to modify the 'hit and run' targeting vector. This was altered to increase the probability of retention of the ΔI506/7 mutation in an excision event, thereby reducing the number of 'true run' clones which must be obtained. The vector was enlarged by the addition of an extra 2.5 kb of homology to the target site, thereby increasing the probability that the mutation would be retained from 17% to 42% of 'run' clones.

This modified vector was targeted to *Cftr* at the high frequency of 14%, generating four targeted clones. One of these 'hit' clones had retained the ΔI506/7 mutation upon integration into the target site, and is therefore suitable for use in the next stage, selection for a 'run' event. As this vector would still be subject to methylation-induced inactivation of the Hygromycin-*tk* fusion gene, the procedure of screening clones for methylation should be utilised.

Therefore future experiments which utilise the HR6 'hit' clone which has retained the mutation for selection of the 'run' event, and which screen the resulting Gancyclovir resistant clones by the modified procedure described in this chapter, should identify more 'true run' clones which have a higher probability of retaining the ΔI506/7 mutation as the only alteration to *Cftr*.

CHAPTER 10
CONCLUDING REMARKS

10.1 THESIS SUMMARY

The aim of this thesis was to create mouse models for cystic fibrosis which carried subtle CF-associated mutations as the only alteration to the CF gene. This was to be achieved by the introduction of precise mutations into ES cells *in vitro* through use of the 'hit and run' gene targeting strategy. Mice could then be generated from such targeted cells and used to study the mutant phenotype.

The first stage of the 'hit and run' targeting strategy was achieved with relative ease. The 'hit and run' vector was targeted to exon 10 of *Cftr* at a high efficiency, generating a number of 'hit' clones which had retained the mutation upon integration, and which were therefore suitable for taking forward into the next stage. The results of these targeting experiments, when combined with data from other experiments also conducted in our laboratory, yielded important information regarding factors which influence the targeting efficiency at this non-selectable locus. In addition, this data also supported the theory that insertion vectors target more reliably and at a higher frequency than replacement vectors (Dickinson *et al.* 1992). These observations were of general interest, as previous reports had all been conducted at the directly-selectable *hprt* locus (Thomas and Capecchi 1987., Hasty *et al.* 1991., Deng and Capecchi 1992).

Selection for the 'run' event of this targeting strategy proved extremely problematic. Clones were found to survive the negative selection with the targeting vector intact, and had become Gancyclovir resistant through a mechanism other than the desired vector excision. This phenomenon was observed in many experiments and occurred at a much higher frequency.

In order to overcome the problem of 'false positive' clones with the targeting vector surviving the negative selection, the mechanism by which this occurred was investigated. Analysis showed that the Hygromycin-*tk* fusion gene of 'run' clones which had been selected in Gancyclovir contained methylated sequences which were absent in the non-selected 'hit' clone from which they were derived. This indicated

that methylation-induced inactivation of the fusion gene was responsible for the loss of Gancyclovir sensitivity in these 'run' clones. The presence of methylated sequences flanking the target site in ES cells and different mouse strains implied that these sequences were normally methylated, and that spread of methylation from this region into the targeting vector was responsible for the loss of *tk* gene activity.

Once the mechanism of Hygromycin-*tk* gene inactivation had been identified, a strategy was devised to identify clones which had become Gancyclovir resistant through this mechanism and eliminate them at an early stage of analysis. By taking advantage of the reversible nature of methylation-induced gene inactivation, clones which had become Gancyclovir resistant through this mechanism were identified by their ability to give rise to Hygromycin resistant cells. Cells which had become Gancyclovir resistant through the desired mechanism of vector excision were not able to reverse their phenotype and remained sensitive to Hygromycin. This strategy was successful in identifying four clones out of the one thousand screened, which had excised the targeting vector. Unfortunately all four of these clones were found to have excised the targeting vector and the mutant exon 10 to revert back to a wildtype status. This outcome was unfortunate but not entirely unexpected as the distribution of the duplicated genomic sequences was such that retention of the mutation might only be expected in one out of six excision events.

To improve the probability of retaining the mutation in subsequent 'run' experiments, the 'hit and run' targeting vector was modified by Dr Paul Dickinson to increase the length of the genomic region carried by the vector. This increased the region of homology available for the recombination event which would result in retention of the mutation, and therefore increased the probability of obtaining a 'true run' clone which has excised the vector and retained the mutation. This vector was successfully targeted to exon 10 of *Cftr*, and generated one clone which had retained the mutation upon integration and which was therefore suitable for use in the 'run' stage. Screening of the resulting 'run' clones by the same modified procedure should

allow the identification of clones which have excised the vector and retained the mutation as the only alteration to *Cftr*.

10.2 DISCUSSION

This project was unsuccessful in introducing subtle mutations into *Cftr* in ES cells using the 'hit and run' targeting strategy. This was due to the difficulty encountered in identifying 'run' clones which had excised the vector from the large number of 'false positive' clones which had become Gancyclovir resistant through the more frequent mechanism of methylation-induced inactivation of the selection genes.

The high frequency at which methylation inactivated the negative selection gene in these experiments is likely to be a consequence of the presence of methylated sequences at the target site, and therefore a property of the target region. This propensity for methylation makes the use of negative selection genes at this locus very difficult. A number of different groups have reported their intention to introduce the ΔF508 mutation into exon 10 of ES cells, using either the 'hit and run' targeting strategy (van Doorninck *et al.* 1993), or the double replacement strategy which also utilises a negative selection step (Ratcliff *et al.* 1992., Koller *et al.* 1991). In light of the findings of this investigation, it is perhaps not surprising therefore, that there are currently no published reports of this being successfully accomplished. Both the 'hit and run' targeting strategy and the double replacement strategy have however, been used successfully to introduce subtle mutations into other loci. The first account of use of the 'hit and run' targeting strategy to introduce mutations into the *hprt* and *Hox 2.6* locus reported that 38% of clones containing the vector survived the negative selection (Hasty *et al.* 1991., Ramirez-Solis *et al.* 1993), and there have also been reports of this phenomenon at other loci. Wu *et al.* (1994) reported that 32% of the clones surviving the negative selection still contained the vector when targeting the murine type I collagen gene using the 'hit and run' targeting strategy, and a lower figure of 22% was reported by Valancius and Smithies (1991) when targeting *hprt* by the same strategy. The same problem has been reported for the second stage of the double replacement strategy which also utilises a negative selection step.

Askew *et al.* (1993) found that 92% of Gancyclovir resistant clones were 'false positives' when targeting an ATPase locus, and Wu *et al.* (1994) reported that 40% still contained the vector when targeting the murine type I collagen gene by the double replacement technique. Although the mechanism giving rise to this loss of sensitivity to the negative selection was not investigated in these cases, it is probable that methylation induced gene inactivation was involved. In all cases, the loss of sensitivity to the negative selection was not accompanied by any detectable alteration to the targeted allele, however, no methylation sensitive restriction enzymes were used to characterise the targeting event. Therefore it seems likely that the presence and extent of the phenomenon of inactivation of negative selection genes when inserted into genomic sequences is highly dependant upon the nature of the target site. The difficulties described previously when introducing subtle mutations into exon 10 by use of techniques which incorporate a negative selection step is therefore probably a reflection of the greater susceptibility of sequences integrated into the region around exon 10 of *Cftr* to become methylated.

Although this project was unsuccessful in generating mouse models for CF which bear subtle CF-associated mutations, it has identified a problem area and led to a modified selection protocol which should facilitate success. The demonstration that clones which have excised the targeting vector can be obtained if a method of reducing the high background of 'false positives' is utilised, has shown that excision of the vector does occur but at a much lower frequency than inactivation of the fusion gene by methylation. Therefore it should still be possible to obtain clones bearing subtle mutations in *Cftr* by this technique, either by screening Gancyclovir resistant clones for resistance to the positive selectable marker, or by using some other method of identifying and eliminating 'false positive' clones. An alternative approach might be to prevent methylation of the selection genes from occurring. Recently a short sequence of DNA called an *Sp1* element has been identified which appears to protect CpG islands from methylation (Brandeis *et al.* 1994., Macleod *et al.* 1994). Incorporation of this element into a β-globin gene construct prevented the *de novo* methylation seen in constructs lacking this sequence on transfection into ES cells.

This effect was seen within a range of approximately 100 bp either side of the *Sp1* element, and therefore if included into a targeting vector near to the promoter, might be sufficient to keep the promoter methylation-free, thereby preventing inactivation of the negative selection gene.

10.3 OVERALL CONCLUSIONS

The overall conclusions from this thesis must be that the success of targeting a locus is highly dependent upon the nature of each particular locus. This thesis has described an attempt to introduce subtle mutations into exon 10 of *Cftr* in ES cells by a proven technique which has been thwarted by the high frequency of *de novo* methylation of vector sequences integrated into this region. This same effect has been reported to varying degrees at other loci, and therefore the ease of achieving the desired result with this targeting technique is highly locus-specific.

The subsequent modification of the 'hit and run' targeting technique has demonstrated that despite the high frequency of methylation-induced gene inactivation, it is possible to identify clones which have undergone the desired event of vector excision. Therefore by utilising clones targeted with a vector which does not favour mutation loss following vector excision, it is still feasible to obtain cells containing CF-associated mutations as the only alteration to *Cftr* by the 'hit and run' gene targeting technique.

10.4 FUTURE DIRECTIONS

Despite the difficulty experienced in introducing subtle mutations into *Cftr*, the potential value of a mouse model created, justifies perseverance towards this end. Modification of the 'hit and run' strategy, and use of clones targeted with greater regions of homology is an obvious approach, however alternatives must be considered. The double replacement strategy uses a different pathway to achieve the same result and could therefore be utilised. However this also employs a negative

selection step and is likely to be susceptible to the same high level of methylation induced 'false positives'.

The presence of a negative selection gene at the target site in 'hit' clones generated in these targeting experiments enables them to be used as the substrate for replacement type targeting. Replacement of the integrated vector including the Hygromycin-*tk* selection genes could be attempted by transfecting 'hit' clones with homologous genomic sequences which span the insertion, as in the second step of the double replacement targeting strategy. The incoming DNA would carry the mutation to be introduced, and through homologous recombination with the endogenous genomic sequences, would replace the vector and duplicated exon 10 sequences. This event would result in loss of the selection gene and could be selected for using Gancyclovir. Although 'false positive' Gancyclovir resistant clones would also arise in this strategy, the frequency of the replacement targeting event might be sufficiently high to allow targeted clones to be identified.

Targeting with a replacement vector which has the non homologous sequences inserted into an intron is another feasible approach. This strategy has been used by Dr Steven Delaney (University of Queensland, Brisbane) to introduce the G551D mutation into exon 11 of *Cftr* (S. Delaney, personal communication). The presence of an expressed gene in an intron can affect expression, and a 30% reduction in expression of the mutant allele is seen in the G551D mice. Reduced expression of a null allele may be of no consequence, however it is an issue when attempting to precisely model the phenotype of missense mutations of a human genetic disease. This especially holds if phenotype/genotype correlations are to be made, and a possible heterozygote advantage studied. One way to overcome this would be to use a recombinase system such as *lox* and *cre* to remove plasmid sequences following targeting. Although plasmid sequences are removed with this technique, it does not create a precise alteration, as a 34 base pair sequence of a *lox* site remains at the target site. Non-targeting strategies for introducing subtle mutations could also be employed, such as crossing transgenes onto a null background. However, although

this creates a model which expresses only the mutant allele of the transgene, it does not model the mutation in its correct chromosomal context. Therefore, for introduction of many of the CF-associated mutations it is worth pursuing the more technically demanding techniques which introduce precise mutations identical to those of the clinical condition.

In conclusion therefore, I believe that having identified the major problem encountered in the 'hit and run' targeting strategy, the creation of mice bearing specific, clinically-relevant mutations is now a realistic possibility. Unfortunately time did not allow for the necessary modifications developed in this work to be fully utilised to achieve this aim. However, the importance of models for CF justifies further work towards this goal. Their use in resolving genotype/phenotype issues, and evaluation of relocation strategies for mislocalised mutant proteins, will be of immense value to the CF community.

REFERENCES

REFERENCES

Alton E. W. F. W., Middleton P. G., Caplen N. J., Smith S. N., Steel D. M., Munkonge F. M., Jeffery P. K., Geddes D. M., Hart S. L., Williamson R., Fasold K. I., Miller A. D., Dickinson P., Stevenson B. J., McLachlan G., Dorin J. R., and Porteous D. J. (1993) Noninvasive liposome-mediated gene delivery can correct the ion-transport defect in cystic fibrosis mutant mice. *Nat.Gen.* **5**, 312-317.

Anderson D. H. (1938) Cystic fibrosis of the pancreas and it's relation to celiac disease: a clinical and pathological study. *Amer. J. Dis. Child.* **72**, 62.

Anderson M. P. and Welsh M. J. (1991) Calcium and cAMP activate different chloride channels in the apical membrane of normal and cystic fibrosis epithelia. *Proc.Natl.Acad.Sci.USA* **88**, 6003-6007.

Anderson M. P. and Welsh M. J. (1992) Regulation by ATP and ADP of CFTR chloride channels that contain mutant nucleotide-binding domains. *Science* **257**, 1701-1704.

Anderson M. P., Berger H. A., Rich D. P., Gregory R. J., Smith A. E., and Welsh M. J. (1991a) Nucleoside triphosphates are required to open the CFTR chloride channel. *Cell* **67**, 775-784.

Anderson M. P., Gregory R. J., Thompson S., Souza D. W., Paul S., Mulligan R. C., Smith A. E., and Welsh M. J. (1991b) Demonstration that CFTR is a chloride channel by alteration of it's anion selectivity. *Science* **253**, 202-205.

Anderson M. P., Rich D. P., Gregory R. J., Smith A. E., and Welsh M. J. (1991c) Generation of cAMP-activated chloride currents by expression of CFTR. *Science* **251**, 679-682.

Anguiano A., Oates R. D., Amos J. A., Dean M., Gerrard B., Stewart C., Maher T. A., White M. B., and Milunsky A. (1992) Congenital bilateral absence of the vas deferens. A primarily genital form of cystic fibrosis. *JAMA* **267**, 1794-1797.

Antequera F., Boyes J., and Bird A. (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* **62**, 503-514.

Arrand J. R., Myers P. A., and Roberts R. J. (1978) A new restriction endonuclease from Streptomyces albus G. *J. Mol. Biol.* **118**, 127-135.

Askew G. R., Doetschman T., and Lingrel J. B. (1993) Site-directed point mutations in embryonic stem cells: A gene-targeting, tag-and-exchange strategy. *Mol.Cell.Biol.* **13**, 4115-4124.

Barasch J., Kiss B., Prince A., Salman L., Gruenert D., and Al-Awqati Q. (1991) Defective acidification of intracellular organelles in cystic fibrosis. *Nature* **352**, 70-73.

Barile M. F. (1973) Mycoplasmal contamination of cell cultures: Mycoplasma-virus-cell culture interactions, in *Contamination in tissue culture* (Fogh J., ed), pp. 131-162. Academic Press, .

Baringa M. (1992) Novel function discovered for the cystic fibrosis gene. *Science* **256**, 444-445.

Baringa M. (1994) Knockout mice: Round two. *Science* **265**, 26-28.

Bear C. E., Duguay F., Naismith A. L., Kartner N., Hanrahan J. W., and Riordan J. R. (1991) Cl- Channel Activity in Xenopus Oocytes Expressing the Cystic fibrosis Gene. *J.Biol.Chem.* **266**, 19142-19145.

Bear C. E., Canhui L., Kartner N., Bridges R. J., Jensen T. J., Ramjeesingh M., and Riordan J. R. (1992) Purification and functional reconstitution of the cystic fibrosis transmembrane regulator (CFTR). *Cell* **68**, 809-818.

Beaudet A., Bocock A., and Buchwald M. (1986) Linkage of cystic fibrosis to two tightly linked DNA markers: joint report from a collaborative study. *Am.J.Hum.Genet.* **39**, 681-693.

Beaudet A. L. (1990) Invited editorial: carrier screening for cystic fibrosis. *Am.J.Hum.Genet.* **47**, 603-605.

Becker P. B., Ruppert S., and Schutz G. (1987) Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell* **51**, 435-443.

Berger H. A., Anderson M. P., Gregory R. J., Thompson S., Howard P. W., Maurer R. A., Mulligan R., Smith A. E., and Welsh M. J. (1991) Identification and regulation of the cystic fibrosis transmembrane conductance regulator-generated chloride channel. *J.Clin.Invest.* **88**, 1422-1431.

Berger W., Hein J., and Gedschold J. (1987) Crossovers in two German cystic fibrosis families determine probe order for MET 7C22 and XV-2c/CS.7. *Hum.Genet.* **77**, 197-199.

Berinstein N., Pennell N., Ottaway C. A., and Shulman M. J. (1992) Gene replacement with one-sided homologous recombination. *Mol.Cell.Biol.* **12**, 360-367.

Bertranpetit J. and Cavalli-Sforza L. L. (1991) A genetic history of the Iberian peninsula. *Ann. Hum. Genet.* **55**, 51-67.

Bird A. (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213.

Bird A. (1992) The essentials of DNA methylation. *Cell* **70**, 5-8.

Birnboim H. C. and Doly J. (1979) A rapid alkaline extraction method for screening recombinant plasmid DNA. *Nucl.Acids.Res.* **7**, 1513-1523.

Boat T. F., Welsh M. J., and Beaudet A. L. (1989) Cystic fibrosis, in *The metabolic basis of inherited disease, 6th ed.* (Scriver C. R., Beaudet A. L., Sly W. S., and Valle D., eds), pp. 2649-2680. McGraw-Hill Inc, New York.

Boyes J. and Bird A. (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123-1134.

Bradbury N. A., Jilling T., Berta G., Sorscher E. J., Bridges R. J., and Kirk K. L. (1992) Regulation of plasma membrane recycling by CFTR. *Science* **256**, 530-532.

Bradley A. (1987) Production and analysis of chimaeric mice, in *Teratocarcinomas and embryonic stem cells a practical approach, 1st ed.* (Robertson E. J., ed), pp. 113-151. IRL Press Ltd, Oxford.

Bradley A., Evans M., Kaufman M. H., and Robertson E. (1984) Formation of germ-line chimaeras from embryo-derived tetrocarcinoma cell lines. *Nature* **309**, 255-256.

Braekeleer M. D. and Daigneault J. (1992) Spatial distribution of the DF508 mutation in cystic fibrosis: a review. *Hum.Biol.* **64**, 167-174.

Brandeis M., Frank D., Keshet I., Siegfried Z., Mendelsohn M., Nemes A., Temper V., Razin A., and Cedar H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**, 435-438.

Braun R. E., Lo D., Pinkert C. A., Widera G., Flavell R. A., Palmiter R. D., and Brinster R. L. (1990) Infertility in male transgenic mice: disruption of sperm development by HSV-tk expression in postmeiotic germ cells. *Biol. Reprod.* **43**, 684-693.

Bullock W. O., Fernandez J. M., and Short J. M. (1987) XL1-Blue - a high efficiency plasmid transforming recA Escherichia coli strain with beta-galactosidase selection. *Biotechniques* **5**, 376.

Buschausen G., Wittig B., Graessmann M., and Graessmann A. (1987) Chromatin structure is required to block transcription of the methylated Herpes Simplex Virus thymidine kinase gene. *Proc.Natl.Acad.Sci.USA* **84**, 1177-1181.

Capecchi M. (1989) The new mouse genetics: altering the genome by gene targeting. *TIG* **5**, 70-76.

Capecchi M. R. (1989) Altering the genome by homologous recombination. *Science* **244**, 1288-1292.

Casals T., Vazquez C., Lazaro C., Girbau E., Gimenez F. J., and Estivill X. (1992) Cystic fibrosis in the Basque country: high frequency of mutation deltaF508 in patients of Basque origin. *Am.J.Hum.Genet.* **50**, 404-410.

Cheadle J., Myring J., Al-Jader L., and Meredith L. (1992) Mutation analysis of 184 cystic fibrosis families in Wales. *J.Med.Genet.* **29**, 642-646.

Chen T. R. (1977) In situ detection of mycoplasma contamination in cell cultures by fluorescent Hoechst 33258 stain. *Exp.Cell Res.* **104**, 255-262.

Cheng S. H., Gregory R. J., Marshall J., Paul S., Souza D. W., White G. A., O'Riordan C. R., and Smith A. E. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **63**, 827-834.

Cheng S. H., Rich D. P., Marshall J., Gregory R. J., Welsh M. J., and Smith A. E. (1991) Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. *Cell* **66**, 1027-1036.

Chu G., Hayakawa H., and Berg P. (1987) Electroporation for the efficient transfection of mammalian cells with DNA. *Nucl.Acids.Res.* **15**, 1311-1326.

Clarke L. L., Grubb B. R., Gabriel S. E., Smithies O., Koller B. H., and Boucher R. C. (1992) Defective epithelial chloride transport in a gene-targeted mouse model of cystic fibrosis. *Science* **257**, 1125-1128.

Clarke L. L., Grubb B. R., Yankaskas J. R., Cotton C. U., McKenzie A., and Boucher R. C. (1994) Relationship of a non-cystic fibrosis transmembrane conductance regulator-mediated chloride conductance to organ-level disease in CFTR (-/-) mice. *Proc.Natl.Acad.Sci.USA* **91**, 479-483.

Cliff W. H. and Frizzell R. A. (1990) Separate Cl- conductances activated by cAMP and Ca2+ in Cl-secreting epithelial cells. *Proc.Natl.Acad.Sci.USA* **87**, 4956-4960.

Cohen L. F., Di Sant'Agnese P. A., and Friedlander J. (1980) Cystic fibrosis and pregnancy. A national survey. *Lancet* **2**, 842-844.

Cohn J. A., Melhus O., Page L. J., Dittrich K. L., and Vigna S. R. (1991) CFTR: development of high-affinity antibodies and localization in sweat gland. *Biochem.Biophys.Res.Comm.* **181**, 36-43.

Colledge W. H., Ratcliff R., Foster D., Williamson R., and Evans M. J. (1992) Cystic fibrosis mouse with intestinal obstruction. *Lancet* **340**, 680.

Collins F. S., Drumm M. L., Cole J. L., Lockwood W. K., Vandewoude G. F., and Iannuzzi M. C. (1987) Construction of a general human-chromosome jumping library, with application to cystic fibrosis. *Science* **235**, 1046-1049.

Cooper G. E., Khattar N. H., Bishop P. L., and Turker M. S. (1992) At least two distinct epigenetic mechanisms are correlated with high frequency 'switching' for APRT phenotypic expression in mouse embryonal carcinoma stem cells. *Som. Cell and Mol. Genet.* **18**, 215-225.

Cormack B. (1994) Mutagenesis by the polymerase chain reaction, in *Current protocols in molecular biology Vol. 1* (Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A., and Struhl K., eds), p. 8.5.5. Greene publishing associates and John Wiley and Sons Inc, .

Coutelle C., Bruckner R., Grade K., Behrens F., Gedschold J., Hein J., Szibor R., Bauer I., Brock J., Graupner I., Urner U., and Leucht B. (1992) Prevalence of cystic fibrosis mutations in the east German population. *Hum.Mut.* **1**, 109-112.

Crawford I., Maloney P. C., Zeitlin P. L., Guggino.WB, Hyde S. C., Turley H., Gatter K. C., Harris A., and Higgins C. F. (1991) Immunocytochemical localization of the cystic fibrosis gene product CFTR. *Proc.Natl.Acad.Sci.USA* **88**, 9262-9266.

Culver K. W., Ram Z., Wallbridge S., Ishii H., Oldfield E. H., and Blaese R. M. (1992) In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors. *Science* **256**, 1550-1552.

Cutting G. R., Kasch L. M., Rosenstein B. J., Tsui L.-C., Kazazian H. H., and Antonarakis S. E. (1990) Two patients with cystic fibrosis, nonsense mutations in each cystic fibrosis gene, and mild pulmonary disease. *New Eng.J.Med.* **323**, 1685-1689.

Dalemans W., Barbry P., Champigny G., Jallet S., Dott K., Dreyer D., Crystal R. G., Pavirani A., Lecocq J., and Lazunski M. (1991) Altered chloride ion channel kinetics associated with the deltaF508 cystic fibrosis mutation. *Nature* **354**, 526-528.

Dalemans W., Hinnrasky J., Slos P., Dreyer D., Fuchey C., Pavirani A., and Puchelle E. (1992) Immunocytochemical analysis reveals differences between the subcellular localization of normal and deltaPhe508 recombinant cystic fibrosis transmembrane conductance regulator. *Exp.Cell Res.* **201**, 235-240.

Davies J., von-Ahsen U., and Schroeder R. (1993), in *The RNA World* (Gesteland R. F. and Atkins J. F., eds), pp. 185-204. Cold spring harbor lab press, .

Davis A. C., Wims M., and Bradley A. (1992) Investigation of coelectroporation as a method for introducing small mutations into embryonic stem cells. *Mol.Cell.Biol.* **12**, 2769-2776.

Dean M., White M. B., Amos J., Gerrard B., Stewart C., Khaw K. T., and Leppert M. (1990) Multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients. *Cell* **61**, 863-870.

Decruyenaere M., Evers-Kiebooms G., Denayer L., and Berghe van den H. (1992) Cystic fibrosis: community knowledge and attitudes towards carrier screening and prenatal diagnosis. *Clin. Genet.* **41**, 189-196.

Deng C. and Capecchi M. R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol.Cell.Biol.* **12**, 3365-3371.

Deng C., Thomas K. R., and Capecchi M. R. (1993) Location of crossovers during gene targeting with insertion and replacement vectors. *Mol.Cell.Biol.* **13**, 2134-2140.

Denning C. R., Sommers S. C., and Quigley H. J. J. (1968) Infertility in male patients with cystic fibrosis. *Pediatrics* **41**, 7-17.

Denning G. M., Anderson M. P., Amara J. F., Marshall J., Smith A. E., and Welsh M. J. (1992a) Processing of the cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* **358**, 761-764.

Denning G. M., Ostegaard L. S., Cheng S. H., Smith A. E., and Welsh M. J. (1992b) Localization of cystic fibrosis transmembrane conductance regulator in chloride secretory epithelia. *J.Clin.Invest.* **89**, 339-349.

Deursen van J. and Wieringa B. (1992) Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. *Nucl.Acids.Res.* **20**, 3815-3820.

Diamond G., Scanlin T. F., Zasloff M. A., and Bevins C. L. (1991) A cross-species analysis of the cystic fibrosis transmembrane conductance regulator. *J.Biol.Chem.* **266**, 22761-22769.

Dickinson P., Kimber W. L., Kilanowski F. M., Stevenson B. J., Porteous D. J., and Dorin J. R. (1993) High frequency gene targeting using insertional vectors. *Hum.Mol.Gen.* **2**, 1299-1302.

Di Sant'Agnese P. A., Darling R. C., Perera G. A., and Shea E. (1953) Abnormal electrolytic composistion of sweat in cystic fibrosis of the pancreas. Clinical significance and relationship of the disease. *Pediatrics* **12**, 549.

Docherty J. G., Zaki A., Coutts J. A. P., Evans T. J., and Carachi R. (1992) Meconium ileus: a review 1972-1990. *Br.J.Surg.* **79**, 571-573.

Dodge J. A. (1991) Determinants of pancreatic and pulmonary status of cystic fibrosis. *Lancet* **337**, 54-55.

Donehower L. A., Harvey M., Slagle B. L., McArthur M. J., Montgomery C. A., Butel J. S., and Bradley A. (1992) Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215-221.

Dorin J. R. and Porteous D. J. (1991) Cystic fibrosis - the way forward from the gene. *TIBTECH* **9**, 48-52.

Dorin J. R., Dickinson P., Alton E. W. F. W., Smith S. N., Geddes D. M., Stevenson B. J., Kimber W. L., Fleming S., Clarke A. R., Hooper M. L., Anderson L., Beddington R. S. P., and Porteous D. J. (1992a) Cystic fibrosis in the mouse by targeted insertional mutagenesis. *Nature* **359**, 211-215.

Dorin J. R., Dickinson P., Emslie E., Clarke A. R., Dobbie L., Hooper M. L., Halford S., Wainwright B. J., and Porteous D. J. (1992b) Successful targeting of the mouse cystic fibrosis transmembrane conductance regulator gene in embryonal stem cells. *Transgen.Res.* **1**, 101-105.

Dorin J. R., Stevenson B. R., Fleming S., Alton E. W. F. W., Dickinson P., and Porteous D. J. (1994) Long term survival of the exon 10 insertional cystic fibrosis mutant mouse is a consequence of low level residual wild-type Cftr gene expression. *Mamm. Gen.* **5**, 465-472.

Dorninck J. H. van, Peters R., Verbeek E., and Scholte R. J. (1993) Towards a CF mouse model with a specific mutation in the CFTR gene. *Pediatr.Pul.* **9**, 221-222.

Drumm M. L., Wilkinson D. J., Smit L. S., Worrell R. T., Strong T. V., Frizzell R. A., Dawson D. C., and Collins F. S. (1991) Chloride conductance expressed by ΔF508 and other mutant CFTRs in Xenopus oocytes. *Science* **254**, 1797-1799.

Dulhanty A. M. and Riordan J. R. (1994) Phosphorlation by cAMP-dependant protein kinase causes a conformational change in the R domain of the cystic fibrosis transmembrane conductance regulator. *biochem* **33**, 4072-4079.

Eden S. and Cedar H. (1994) Role of DNA methylation in the regulation of transcription. *Curr. Opin. in Genet. and Dev.* **4**, 255-259.

Egan M., Flotte T., Afione S., Solow R., Zeitlin P. L., Carter B. J., and Guggino W. B. (1992) Defective regulation of outwardly rectifying Cl- channels by protein kinase A corrected by insertion of CFTR. *Nature* **358**, 581-584.

Eiberg H., Mohr J., Schmiegelow K., Neilson L. S., and Williamson R. (1985) Linkage relationships of paraoxonase (PON) with other markers: identification of PON-cystic fibrosis synteny. *Clin. Genet.* **28**, 265-271.

Elborn J. S. and Shale D. J. (1990) Lung injury in cystic fibrosis. *Thorax* **45**, 970-973.

Elborn J. S., Shale D. J., and Britton J. R. (1991) Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax* **91?**, 881-885.

Estivill X., Farrell M., and Scrambler P. J. (1987) A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature* **326**, 840-845.

Evans M. J. and Kaufman M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156.

Fanconi G., Uehlinger E., and Knauer C. (1936) Das coeliakiesyndrom be: angeborener zysticher pankreas fibromatose und bronkiekasien. *Wien Med. Wochenschr* **86**, 753.

Farber S. (1945) Some organic digestive disturbances in early life. *J. Michigan Med. Soc.* **44**, 587.

Farrell M., Wainwright B. J., and Feldman G. L. (1988) Recombination between IRP and cystic fibrosis. *Am.J.Hum.Genet.* **43**, 741-775.

Feinberg A. P. and Vogelstein B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6-13.

Ferec C., Audrezet M. P., Mercier B., Guillermit H., Moullier P., Quere I., and Verlingue C. (1992) Detection of over 98% cystic fibrosis mutations in a Celtic population. *Nat.Gen.* **1**, 188-191.

Ferec C., Verlingue C., Guillermit H., Quere I., Raguenes O., Feigelson J., Audrezey M. P., Moullier P., and Mercier B. (1993) Genotype analysis of adult cystic fibrosis patients. *Hum.Mol.Gen.* **2**, 1557-1560.

Folger K. R., Wong E. A., and Wahl G. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol.Cell.Biol.* **2**, 1372-1387.

Frizzell R. A. and Cliff W. H. (1991) Back to the chloride channel. *Nature* **350**, 277-278.

Frizzell R. A., Rechkemmer G., and Shoemaker R. L. (1986) Altered regulation of airway epithelial cell chloride channels in cystic fibrosis. *Science* **233**, 558-560.

Gabriel S. E., Clarke L. L., Boucher R. C., and Stutts M. J. (1993) CFTR and outwardly rectifying chloride channels are distinct proteins with a regulatory relationship. *Nature* **363**, 263-266.

Gabriel S. E., Brigman K. N., Koller B. H., Boucher R. C., and Stutts M. J. (1994) Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science* **266**, 107-109.

Gasparini P., Nunes V., Dogini M., and Estivill X. (1991) High conservation of sequences involved in cystic fibrosis mutations in five mammalian species. *Genomics* **10**, 1070-1072.

Gibson L. E. and Cooke R. E. (1959) A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilising pilocarpine by iontophoresis. *Pediatrics* **23**, 545-549.

Gilbert F. (1989) Is population screening for cystic fibrosis appropriate now? *Am.J.Hum.Genet.* **46**, 394-395.

Gill D. R., Hyde S. C., Higgins C. F., Valverde M. A., Mintenig G. M., and Sepulveda F. V. (1992) Separation of drug transport and chloride channel functions of the human multidrug resistance P-glycoprotein. *Cell* **71**, 23-32.

Gordon J. W., Scangos G. A., Plotkin D. J., Barbosa J. A., and Ruddle F. H. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc.Natl.Acad.Sci.USA* **77**, 7380-7384.

Govan J. R. W., Brown P. H., Maddison J., Doherty C. J., Nelson J. W., Dodd M., Greening A. P., and Webb A. K. (1993) Evidence for transmission of Pseudomonas-cepacia by social contact in cystic fibrosis. *Lancet* **342**, 15-19.

Graessmann A., Sandberg G., Guhl E., and Graessmann M. (1994) Methylation of single sites within the Herpes simplex virus tk coding region and the Simian virus 40 T-antigen intron causes gene inactivation. *Mol.Cell.Biol.* **14**, 2004-2010.

Graves K. H. and Moreadith R. W. (1993) Derivation and characterisation of putative pluripotential embryonic stem cells from preimplantation rabbit embryos. *Mol. Reprod. Dev.* **36**, 424-433.

Gray M. A., Winpenny J. P., Porteous D. J., Dorin J. R., and Argent B. E. (1994) CFTR and calcium-activated chloride currents in pancreatic duct cells of a transgenic CF mouse. *Am.J.Physiol.* **266**, 213-221.

Gregory R. J., Rich D. P., Cheng S. H., Souza D. W., Paul S., Manavalan P., Anderson M. P., Welsh M. J., and Smith A. E. (1991) Maturation and function of cystic fibrosis transmembrane conductance regulator varients bearing mutations in putative nucleotide-binding domains 1 and 2. *Mol.Cell.Biol.* **11**, 3886-3893.

Grubb B. R., Pickles R. J., Ye H., Yankaskas J. R., Vick R. N., Engelhardt J. F., Wilson J. M., Johnson L. G., and Boucher R. C. (1994) Inefficient gene transfer by adenovirus vector to cystic fibrosis airway epithelia of mice and humans. *Nature* **371**, 802-806.

Hamosh A., King T. M., Rosenstein B. J., Corey M., Levison H., Durie P., Tsui L.-C., McIntosh I., Keston M., Brock D. J. H., Macek M. J., Zemkova D., Krasnicanova H., Vavrova V., Macek M. S., Golder N., Schwarz M. J., Super M., Watson E. K., Williams C., Bush A., O'Mahoney S. M., Humphries P., DeArce M. A., Reis A., Burger I., Stuhrmann M., Schmidtke J., Wulbrand U., Dork T., Tummler B., and Cutting G. R. (1992) Cystic fibrosis patients bearing both the common missense mutation Gly-Asp at codon 551 and the deltaF508 mutation are clinically indistinguishable from deltaF508 homozygotes except for decreased risk of meconium ileus. *Am.J.Hum.Genet.* **51**, 245-250.

Hanahan D. (1983) Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557-580.

Handyside A. H., O'Neill G. T., Jones M., and Hooper M. L. (1989) Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line. *Roux's Arch. Devel. Biol.* **198**, 48-55.

Hartman J., Huang Z., Rado T. A., Peng S., Jilling T., Muccio D. D., and Sorscher E. J. (1992) Recombinant synthesis, purification, and nucleotide binding characteristics of the first nucleotide binding domain of the cystic fibrosis gene product. *J.Biol.Chem.* **267**, 6455-6458.

Hasegawa H., Skach W., Baker O., Calayag M. C., Lingappa V., and Verkman A. S. (1992) A multifunctional aqueous channel formed by CFTR. *Science* **258**, 1477-1479.

Hastings P. J., McGill C., Shafer B., and Strathern J. N. (1993) Ends-In vs ends-out recombination in yeast. *Genetics* **135**, 973-980.

Hasty P. and Bradley A. (1993) Gene targeting vectors for mammalian cells, in *Gene Targeting, 1st ed.* (Joyner A. L., ed), pp. 1-30. IRL Press, New York.

Hasty P., Ramirez-Solis R., Krumlauf R., and Bradley A. (1991a) Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. *Nature* **350**, 243-246.

Hasty P., Rivera-Perez J., and Bradley A. (1991b) The length of homology required for gene targeting in embryonic stem cells. *Mol.Cell.Biol.* **11**, 5586-5991.

Hasty P., Rivera-Perez J., Chang C., and Bradley A. (1991c) Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Mol.Cell.Biol.* **11**, 4509-4517.

Hasty P., Rivera-Perez J., and Bradley A. (1992) The role and fate of DNA ends for homologus recombination in embryonic stem cells. *Mol.Cell.Biol.* **12**, 1464-2474.

Higgins C. (1989) Protein joins transport family. *Nature* **341**, 103.

Higgins C. F. and Hyde S. C. (1991) Channelling our thoughts. *Nature* **352**, 194-195.

Highsmith W. E., Burch L. H., Zhou Z., Olsen J. C., Boat T. E., Spock A., Gorvoy J. D., Quittell L., Friedman K. J., Silverman L. M., Boucher R. C., and Knowles M. R. (1994) A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. *New Eng.J.Med.* **331**, 974-980.

Hinnen A., Hicks J. D., and Fink G. R. (1978) Transformation of yeast. *Proc.Natl.Acad.Sci.USA* **75**, 1929-1933.

Hooper M., Hardy K., Handyside A., Hunter S., and Monk M. (1987) HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292-295.

Hooper M. L. (1987) Isolation of genetic varients and fusion hybrids from embryonal carcinoma cell lines, in *Teratocarcinomas and embryonic stem cells* (Robertson E. J., ed), pp. 51-71. IRL Press, Oxford.

Hwang T.-C., Lu L., Zeitlin P. L., Gruenert D. C., Huganir R., and Guggino W. B. (1989) Cl channels in CF: Lack of activation by protein kinase C and cAMP-dependent protein kinase. *Science* **244**, 1351-1353.

Hyde S. C., Emsley P., Hartshorn M. J., Mimmack M. M., Gileadi U., Pearce S. R., Gallagher M. P., Gill D. R., Hubbard R. E., and Higgins C. F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis,multidrug resistance and bacterial transport. *Nature* **346**, 362-365.

Hyde S. C., Gill D. R., Higgins C. F., and Trezise A. E. O. (1993) Correction of the ion transport defect in cystic fibrosis transgenic mice by gene therapy. *Nature* **362**, 250-255.

Iannaccone P. M., Greg T. U., Garton R. L., Caplice M. D., and Brenin D. R. (1994) Pluripotent embryonic stem cells from the rat are capable of producing chimeras. *Dev. Biol.* **163**, 288-292.

Jequier A. M., Ansell I. D., and Bullimore N. J. (1985) Congenital absence of the vasa deferentia presenting with infertility. *J. Androl.* **6**, 15-19.

Jetten A. M., Yankaskas J. R., Stutts M. J., Willumsen N. J., and Boucher R. C. (1989) Persistence of abnormal chloride conductance regulation in transformed cystic fibrosis epithelia. *Science* **244**, 1472-1475.

Jiang C., Finkbeiner W. E., Widdicombe J. H., McCray P. B., and Miller S. S. (1993) Altered fluid transport across airway epithelium in Cystic fibrosis. *Science* **262**, 424-427.

Johansen H. K., Nir M., Hoiby N., Koch C., and Schwartz M. (1991) Severity of cystic fibrosis in patients homozygous and heterozygous for deltaF508 mutation. *Lancet* **337**, 631-634.

Kafri T., Ariel M., Brandeis M., Shemer R., Urven L., McCarrey J., Cedar H., and Razin A. (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes and Dev.* **6**, 705-714.

Kaplan E., Shwachman H., Perlmutter A. D., Rule A., Khaw K.-T., and Holsclaw D. S. (1968) Reproductive failure in males with cystic fibrosis. *New Eng.J.Med.* **279**, 65-69.

Kartner N., Hanrahan J. W., Jensen T. J., Naismith A. L., Sun S., Ackerley C. A., Reyes E. F., Tsui L.-C., Rommens J. M., Bear C. E., and Riordan J. R. (1991) Expression of the cystic fibrosis gene in non-epithelial invertebrate cells produces a regulated anion conductance. *Cell* **64**, 681-691.

Kartner N., Augustinas O., Jensen T. J., Naismith A. L., and Riordan J. R. (1992) Mislocalization of deltaF508 CFTR in cystic fibrosis sweat gland. *Nat.Gen.* **1**, 321-327.

Kass S. U., Goddard J. P., and Adams R. L. P. (1993) Inactive chromatin spreads from a focus of methylation. *Mol.Cell.Biol.* **13**, 7372-7379.

Kelley K. A., Stamm S., and Kozak C. A. (1992) Expression and chromosome localization of the murine cystic fibrosis transmembrane conductance regulator. *Genomics* **13**, 381-388.

Kerem B., Rommens J. M., Buchanan J. A., Markiewicz D., Cox T. K., Chakravarti A., Buchwald M., and Tsui L.-C. (1989) Identification of the cystic fibrosis gene: Genetic analysis. *Science* **245**, 1073-1080.

Kerem B., Zielenski J., Markiewicz D., Bozon D., Gazit E., Yahav J., Kennedy D., Riordan J. R., Collins F. S., Rommens J. M., and Tsui L.-C. (1990) Identification of mutations in regions corresponding to the two putative nucleotide (ATP)-binding folds of the cystic fibrosis gene. *Proc.Natl.Acad.Sci.USA* **87**, 8447-8451.

Kerem E., Corey M., Kerem B.-S., Rommens J., Markiewicz D., Levison H., Tsui L.-C., and Durie P. (1990) The relationship between genotype and phenotype in cystic fibrosis-analysis of the most common mutation (deltaF508). *New Eng.J.Med.* **323**, 1517-1522.

Keshet I., Lieman-Hurwitz J., and Cedar H. (1986) DNA methylation affects the function of active chromatin. *Cell* **44**, 535-543.

Kiesewetter S., Macek M., Curristin S., Chu C., Graham C., Shrimpton A. E., Dearce M., Zielenski J., Mickle J., Amos J., Highsmith W. E., Shuber A., Witt D., Crystal R. G., and Cutting G. R. (1993) The CFTR mutation R117H produces different phenotpyes depending on the genetice background. *Am.J.Hum.Genet.* **53**, 86.

Knowles M., Gatzy J., and Boucher R. (1981) Increased bioelectric potential difference across respiratory epithelia in cystic fibrosis. *New Eng.J.Med.* **305**, 1489-1495.

Knowles M., Gatzy J., and Boucher R. (1983) Relative ion permeability of normal and cystic fibrosis nasal epithelium. *J.Clin.Invest.* **71**, 1410-1417.

Knowles M. R., Stutts M. J., Spock A., Fischer N., Gatzy J. T., and Boucher R. C. (1983) Abnormal ion permeation through cystic fibrosis respiratory epithelium. *Science* **221**, 1067-1070.

Knowles M. R., Stutts M. J., Yankaskas J. R., Gatzy J. T., and Boucher R. C. (1986) Abnormal respiratory epithelial ion transport in cystic fibrosis. *Clin. Chest Med.* **7**, 285-297.

Knudson A. G., Wayne L., and Hallet W. Y. (1967) On the selective advantage of cystic fibrosis heterozygotes. *Am.J.Hum.Genet.* **19**, 388-392.

Kobayashi K., Knowles M. R., Boucher R. C., O'Brien W. E., and Beaudet A. L. (1990) Benign missense variations in the cystic fibrosis gene. *Am.J.Hum.Genet.* **47**, 611-615.

Koh J., Sferra T. J., and Collins F. S. (1993) Characterisation of the cystic fibrosis transmembrane conductance regulator promoter region. *J.Biol.Chem.* **268**, 15912-15921.

Kolberg R. (1992) Animal models point the way to human clinical trials. *Science* **256**, 772-773.

Koller B. H., Kim H., Latour A. M., Brigman K., Boucher R. C., Scrambler P., Wainwright B., and Smithies O. (1991) Towards an animal model of cystic fibrosis: Targeted disruption of exon 10 of the cystic fibrosis transmembrane conductance regulator gene in embryonic stem cells. *Proc.Natl.Acad.Sci.USA* **88**, 10730-10734.

Koopman P., Gubbay J., Vivian N., Goodfellow P., and Lovell-Badge R. (1991) Male development of chromosomally female mice transgenic for *SRY*. *Nature* **351,** 117-121.

Kotloff R. M., FitzSimmons S. C., and Fiel S. B. (1992) Fertility and pregnancy in patients with cystic fibrosis. *clin. in chest med.* **13**, 623-635.

Kristidis P., Bozon D., Corey M., Markiewicz D., Rommens J., Tsui L.-C., and Durie P. (1992) Genetic determination of exocrine pancreatic function in cystic fibrosis. *Am.J.Hum.Genet.* **50**, 1178-1184.

Kuehn M. R., Bradley A., Robertson E. J., and Evans M. J. (1987) A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326**, 295-298.

Kumar S. and Simons J. P. (1993) The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells. *Nucl.Acids.Res.* **21**, 1541-1548.

Laird P. W., Zijderveld A., Linders K., Rudnicki M., Jaenisch R., and Berns A. (1991) Simplified mammalian DNA isolation procedure. *Nucl.Acids.Res.* **19**, 4293.

Lathe R. and Mullins J. J. (1993) Transgenic animals as models for human-disease-report of an EC study-group. *Transgen.Res.* **2**, 286-299.

Lerer I., Sagi M., Cutting G. R., and Abeliovich D. (1992) Cystic fibrosis mutations deltaF508 and G542X in Jewish patients. *J.Med.Genet.* **29**, 131-133.

Lesch M. and Nyhan W. L. (1964) A familial disorder of uric acid metabolism and central nervous system function. *Am. J. Med.* **36**, 561-570.

Letterio J. J., Geiser A. G., Kulkarni A. B., Roche N. S., Sporn M. B., and Roberts A. B. (1994) Maternal rescue of transforming factor-beta 1 null mice. *Science* **264**, 1936-1938.

Li C., Ramjeesingh M., Reyes E., Jensen T., Chang X., Rommens J. M., and Bear C. E. (1993) The cystic fibrosis mutation (deltaF508) does not influence the chloride channel activity of CFTR. *Nat.Gen.* **3**, 311-316.

Li M., McCann J. D., Liedtke C. M., Nairn A. C., Greengard P., and Welsh M. J. (1988) Cyclic AMP-dependent protein kinase opens chloride channels in normal but not cystic fibrosis airway epithelium. *Nature* **331**, 358-360.

Li M., McCann J. D., Anderson M. P., Clancy J. P., Liedtke C. M., Nairn A. C., Greengard P., and Welsh M. J. (1989) Regulation of chloride channels by protein kinase C in normal and cystic fibrosis airway epithelia. *Science* **244**, 1353-1356.

Lin F. L., Sperle K., and Sternberg N. (1984) Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Mol.Cell.Biol.* **4**, 1020-1034.

Love P. E., Tremblay M. L., and Westphal H. (1992) Targeting of the T-cell receptor chain gene in embryonic stem cells:Strategies for generating multiple mutations in a single gene. *Proc.Natl.Acad.Sci.USA* **89**, 9929-9933.

Lukacs G. L., Chang X.-B., Kartner N., Rotstein O. D., Riordan J. R., and Grinstein S. (1992) The cystic fibrosis transmembrane conductance regulator is present and functional in endosomes. *J.Biol.Chem.* **267**, 14568-14572.

Macleod D., Charlton J., Mullins J., and Bird A. P. (1994) Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes and Dev.* **in press**.

Mansour S. L., Thomas K. R., and Capecchi M. R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-352.

Marshall J., Martin K. A., Picciotti M., Hockfield S., Nairn A. C., and Kaczmarek L. K. (1991) Identification and localization of a dogfish homolog of human cystic fibrosis transmembrane conductance regulator. *J.Biol.Chem.* **266**, 22749-22754.

Marteau T. M., Duijn M. van, and Ellis I. (1992) Effects of genetic screening on perceptions of health: a pilot study. *J.Med.Genet.* **29**, 24-26.

Martin G. R. (1975) Teratocarcinomas as a model system for the study of embryogenesis and neoplasia. *Cell* **5**, 229-243.

Martin G. R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc.Natl.Acad.Sci.USA* **78**, 7634-7638.

Miller S. R. and Schartz R. H. (1992) Attitudes towards genetic testing of Amish, Mennonite, and Hutterite families with cystic fibrosis. *Am. J.Pub. Health* **82**, 236-242.

Mintz B. and Illmensee K. (1975) Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc.Natl.Acad.Sci.USA* **72**, 3585-3589.

Moens C. B., Auerbach A. B., Conlon R. A., Joyner A. L., and Rossant J. (1992) A targeted mutation reveals a role for N-myc in branching morphogenesis in the embryonic mouse lung. *Genes and Dev.* **6**, 691-704.

Morral N., Bertranpetit J., Estivill X., Nunes V., Casals T., Gimenez J., Reis A., VaronMateeva R., Macek M., Kalaydjieva L., Angelicheva D., Dancheva R., Romeo G., Russo M. P., Garnerone S., Restagno G., Ferrari M., Magnani C., and Claustres M. (1994) The origin of the major cystic fibrosis mutation (deltaF508) in European populations. *Nat.Gen.* **7**, 169-175.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G., and Erlich H. (1986) Specific enzymatic amplification of DNA in vitro - The polymerase chain reaction. *CSH Symp. Quant. Biol.* **51**, 263-273.

Nagy A. and Rossant J. (1993) Production of completely ES cell-derived fetuses, in *Gene targeting, 1st ed.* (Joyner A. L., ed), pp. 147-178. IRL Press, New York.

Oates R. D., Honig S., Berger M. J., and Harris D. (1992) Microscopic epididymal sperm aspiration (MESA): a new option for treatment of the obstructive azoospermia associated with cystic fibrosis. *J. Ass. Repro. Genet.* **9**, 36-40.

Okayama H., Curiel D. T., Brantly M. L., Holmes M. D., and Crystal R. G. (1989) Rapid non radioactive detection of mutations in the human genome by allele specific amplification. *J.Lab.Clin.Med.* **114**, 105-109.

O'Neal W. K., Hasty P., McCray Jr P. B., Casey B., Rivera-Perez J., Welsh M. J., Beaudet A. L., and Bradley A. (1993) A severe phenotype in mice with a duplication of exon 3 in the cystic fibrosis locus. *Hum.Mol.Gen.* **2**, 1561-1569.

Orr-Weaver T. L., Szostak J. W., and Rothstein R. J. (1981) Yeast transformation: A model system for the study of recombination. *Proc.Natl.Acad.Sci.USA* **78**, 6354-6358.

Osborne L., Santis G., Schwarz M., Klinger K., Dork T., McIntosh I., Schwartz M., Nunes V., Macek M. J., Reiss J., Highsmith W. E. J., McMahon R., Novelli G., Malik N., Burger J., Anvret M., Wallace A., Williams C., Mathew C., Rozen R., Graham C., Gasparini P., Bal J., Cassiman J. J., Balassopoulou A., Davidow L., Estivill X., Vavrova V., Friedman K. J., Barton D., Dallapiccola B., Stuhrmann M., Beards F., Hill A. J. M., Pignatti P. F., Cuppens H., Angelicheva D., Tummler B., Brock D. J. H., Casals T., Macek M., Schmidtke J., Magee A. C., Bonizzato A., DeBoeck C. D., Kuffardjieva A., Hodson M., and Knight R. A. (1992) Incidence and expression of the cystic fibrosis (CFTR) gene. *Hum.Genet.* **89**, 653-658.

Osborne L. R., Lynch M., Middleton P. G., Alton E. W. F. W., Geddes D. M., Pryor J. P., Hodson M. E., and Santis G. K. (1993) Nasal epithelial ion transport and genetic analysis of infertile men with congenital bilateral absence of the vas deferens. *Hum.Mol.Gen.* **2**, 1605-1609.

Pignatti P. F. (1991) Cystic fibrosis gene mutations and correlation with clinical manifestations. *Path.Biol.* **39**, 582-584.

Porteous D. J. and Alton E. W. F. W. (1993) Cystic fibrosis: Prospects for therapy. *BioEssays* **15**, 485-486.

Porteous D. J. and Dorin J. R. (1993) Gene therapy for cystic fibrosis-where and when? *Hum.Mol.Gen.* **2**, 211-212.

Quinton P. M. (1983) Chloride impermeability in cystic fibrosis. *Nature* **301**, 421-422.

Quinton P. M. and Reddy M. M. (1992) Control of CFTR chloride conductance by ATP levels through non-hydrolytic binding. *Nature* **360**, 79-81.
Ramirez-Solis R., Zeng H., Whiting J., Krumlauf R., and Bradley A. (1993) Hox-4b (Hox-2.6) mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. *Cell* **73**, 279-294.

Ratcliff R., Evans M. J., Doran J., Wainwright B. J., Williamson R., and Colledge W. H. (1992) Disruption of the cystic fibrosis transmembrane conductance regulator gene in embryonic stem cells by gene targeting. *Transgen.Res.* **1**, 177-181.

Ratcliff R., Evans M. J., Cuthbert A. W., MacVinish L. J., Foster D., Anderson J. R., and Colledge W. H. (1993) Production of a severe cystic fibrosis mutation in mice by gene targeting. *Nat.Gen.* **4**, 35-41.

Reid L. H., Gregg R. G., Smithies O., and Koller B. H. (1990) Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. *Proc.Natl.Acad.Sci.USA* **87**, 4299-4303.

Reid L. H., Shesely E. G., Kim H.-S., and Smithies O. (1991) Cotransformation and gene targeting in mouse embryonic stem cells. *Mol.Cell.Biol.* **11**, 2769-2777.

Rich D. P., Anderson M. P., Gregory R. J., Cheng S. H., Paul S., Jefferson D. M., McCann J. D., Klinger K. W., Smith A. E., and Welsh M. J. (1990) Expression of cystic fibrosis transmembrane conductor regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. *Nature* **347**, 358-363.

Rich D. P., Gregory R. J., Anderson M. P., Manavalan P., Smith A. E., and Welsh M. J. (1991) Effect of deleting R domain on CFTR-generated chloride channels. *Science* **253**, 205-207.

Richardson P. S. and Alton E. W. F. W. (1993) Cystic fibrosis transmembrane conductance regulator protein-what isits role in cystic fibrosis? *Eur. Resp. J.* **6**, 160-162.

Riele H. T., Maandag E. R., and Berns A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc.Natl.Acad.Sci.USA* **89**, 5128-5132.

Rigot J.-M., Lafitte J.-J., Dumur V., Gervais R., Manouvrier S., Biserte J., Mazeman E., and Roussel P. (1991) Cystic fibrosis and congenital absence of the vas deferens. *New Eng.J.Med.* **325**, 64-65.

Ringe D. and Petsko G. A. (1990) A transport problem? *Nature* **346**, 312-313.

Riordan J. R., Rommens J. M., Kerem B., Alon N., Rozmahel R., Grzelczak Z., Zielenski J., Lok S., Plavsic N., Chou J., Drumm M., Iannuzzi M. C., Collins F. S., and Tsui L.-C. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066-1073.

Roberts L. (1990) CF screening delayed for awhile, perhaps forever. *Science* **247**, 1296-1297.

Robertson E. J. (1987) Embryo-derived stem cells, in *Teratocarcinomas and embryonic stem cells, 1st ed.* (Robertson E. J., ed), pp. 71-112. IRL Press Ltd, Oxford.

Rodman D. M. and Zamudio S. (1991) The cystic fibrosis heterozygote - advantage in surviving cholera? *Med.Hypoth.* **36**, 253-258.

Rommens J. M., Iannuzzi M. C., Kerem B.-S., Drumm M. L., Melmer G., Dean M., Rozmahel R., Cole J. L., Kennedy D., Hidaka N., Zsiga M., Buchwald M., Riordan J. R., Tsui L.-C., and Collins F. S. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059-1065.

Roth D. B., Porter T. N., and Wilson J. H. (1985) Mechanisms of nonhomologous recombination in mammalian cells. *Mol.Cell.Biol.* **5**, 2599-2607.

Rothstein R. (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast, in *methods in enzymology Vol. 194*, pp. 281-301. Academic press inc, .

Rubnitz J. and Subramani S. (1984) The minimun amount of homology required for homologous recombination in mammalian cells. *Mol.Cell.Biol.* **4**, 2253-2258.

Ryley H. C., Robinson P. G., Yamashiro Y., and Bradley D. M. (1981) Assay os serum immunoreactive trypsin in dried blood and the early detection of cystic fibrosis. *J.Clin.Pathol.* **34**, 906-910.

Saiki R. K., Scharf S., Faloona F., Mullis K., Horn G., Erlich H. A., and Arnheim N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Sambrook J., Fritsch E. F., and Maniatis T. (1989) *Molecular cloning. A laboratory manual, 2nd ed.* Cold Spring Harbor Laboratory Press, .

Santamaria F., Salvatore D., Castiglione O., Raia V., Ritis G. de, and Sebastio G. (1992) Lung involvement, the deltaF508 mutation and DNA haplotype analysis in cystic fibrosis. *Hum.Genet.* **88**, 639-641.

Santis G., Osborne L., Knight R. A., and Hodson M. E. (1990a) Independant genetic determinants of pancreatic and pulmonary status in cystic fibrosis. *Lancet* **336**, 1081-1084.

Santis G., Osborne L., Knight R. A., and Hodson M. E. (1990b) Linked marker haplotypes and the deltaF508 mutation in adults with mild pulmonary disease and cystic fibrosis. *Lancet* **335**, 1426-1429.

Sarkadi B., Bauzon D., Huckle W. R., Earp H. S., Berry A., Suchindran H., Price E. M., Olsen J. C., Boucher R. C., and Scarborough G. A. (1992) Biochemical characterisation of the cystic fibrosis transmembrane conductance regulator in normal and cystic fibrosis epithelial cells. *J.Biol.Chem.* **267**, 2087-2095.

Sawai S., Shimono A., Hanaoka K., and Kondoh H. (1991) Embryonic lethality resulting from disruption of both N-myc alleles in mouse zygotes. *New Biol.* **3**, 861-869.

Scambler P. J., Law H. Y., Williamson R., and Cooper C. S. (1986) Chromosome mediated gene transfer of sic DNA markers linked to the cystic fibrosis locus on human chromosome 7. *Nucl.Acids.Res.* **14**, 7159-7174.

Scherer S. and Davis R. W. (1979) Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc.Natl.Acad.Sci.USA* **76**, 4951-4955.

Schoumacher R. A., Shoemaker R. L., Halm D. R., Tallant E. A., Wallace R. W., and Frizzell R. A. (1987) Phosphorylation fails to activate chloride channels from cystic fibrosis airway cells. *Nature* **330**, 752-754.

Sealey P. G., Whittaker P. A., and Southern E. M. (1985) Removal of repeated sequences from hybridization probes. *Nucl.Acids.Res.* **13**, 1905-1922.

Sheppard D. N., Rich D. P., Ostedgaard L. S., Gregory R. J., Smith A. E., and Welsh M. J. (1993) Mutations in CFTR associated with mild disease form Cl- channels with altered pore properties. *Nature* **362**, 160-164.

Sheppard D. N., Ostedgaard L. S., Rich D. P., and Welsh M. J. (1994) The amino-terminal portion of CFTR forms a regulated Cl- channel. *Cell* **76**, 1091-1098.

Shortle D., Haber J. E., and Botstein D. (1982) Lethal disruption of the yeast actin gene by integrative DNA transformation. *Science* **217**, 371-373.

Shoshani T., Augarten A., Gazit E., Bashan S., Yahav Y., Rivlin Y., Tal A., Seret H., Yaar L., Kerem E., and Kerem B. (1992) Association of a nonsense mutation (W1282X), The most common mutation in the Ashenazi jewish cystic fibrosis patients in Israel, with presentation of severe disease. *Am.J.Hum.Genet.* **50**, 222-228.

Shull M. M., Ormsby I., Kier A. B., Pawlowski S., Diebold R. J., Yin M. Y., Allen R., Sidman C., Proetzel G., Calvin D., Annunziata N., and Doetschman T. (1992) Targeted disruption of mouse transforming growth factor-beta-1 gene results in multifocal inflammatory disease. *Nature* **359**, 693-699.

Smith A. G. and Hooper M. L. (1987) Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev. Biol.* **121**, 1-9.

Smith A. G., Heath J. K., Donaldson D. D., Wong G. G., Moreau J., and Rogers D. (1988) Inhibition of pluripotent embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690.

Smith D. L., Gunnery L. B., Smith E. G., Stableforth D. E., Kaufman M. E., and Pitt T. L. (1993) Epidemic of Pseudomonas cepacia in an adult CF unit:evidence of person-to-person transmission. *J. Clin. Microbiol.* **31**, 3017-3022.

Smith S. N., Alton E. W. F. W., and Geddes D. M. (1992) Ion transport characteristics of the murine trachea and caecum. *Clin.Sci.* **82**, 667-672.

Smith T. A. and Hooper M. L. (1983) Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Exp.Cell Res.* **145**, 458-462.

Smithies O. and Kim H.-S. (1994) Targeted gene duplication and disruption for analyzing quantitive genetic traits in mice. *Proc.Natl.Acad.Sci.USA* **91**, 3612-3615.

Smithies O., Gregg R. G., Boggs S. S., Koralewski M. A., and Kucherlapati R. S. (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230-234.

Snouwaert J. N., Brigman K. K., Latour A. M., Malouf N. N., Boucher R. C., Smithies O., and Koller B. H. (1992) An animal model for cystic fibrosis made by gene targeting. *Science* **257**, 1083-1088.

Solter D. and Skreb N. (1970) Extrauterine growth of mouse egg-cylinders results in malignant teratoma. *Nature* **227**, 503-504.

Southern E. M. (1975) Detection of specific sequences among DNA fragments separated by gel elctrophoresis. *J. Mol. Biol.* **137**, 120-124.

Stacey A., Schnieke A., McWhir J., Cooper J., Colman A., and Melton D. W. (1994) Use of double-replacement gene targeting to replace the murine alpha-lactalbumin gene with its human counterpart in embryonic stem cells and mice. *Mol.Cell.Biol.* **14**, 1009-1016.

Stamatoyannopoulus G. (1974) Problems of screening and counselling in the hemoglobinopathies, in *Birth Defects. Proceedings of the fourth international conference* (Motulsky A. G. and Ebling F. J. G., eds), pp. 268-276. Exerpta Medica, Vienna.

Stutts M. J., Cotton C. U., Yankaskas J. R., Cheng E., Knowles M. R., Gatzy J. T., and Boucher R. C. (1985) Chloride uptake into cultured airway epithelial cells from cystic fibrosis patients and normal individuals. *Proc.Natl.Acad.Sci.USA* **82**, 6677-6681.

Sukoyan M. A., Vatolin S. Y., Golubitsa A. N., Zhelezova A. I., Semenova L. A., and Serov O. L. (1993) Embryonic stem cells derived from morulae, inner cell mass, and blastocysts of mink: Comparisons of their pluripotencies. *Mol. Reprod. Dev.* **36**, 148-158.

Super M. and Schwarz M. J. (1992) Mutations of the cystic fibrosis gene locus within the population of the northwest of England. *Eur. J. Pediatr.* **151**, 108-111.

Super M., Schwarz M. J., and Malone G. (1992) Screening for cystic fibrosis. *Lancet* **340**, 490-491.

Szostak J. W., Orr-Weaver T. L., Rothstein R. J., and Stahl F. W. (1983) The double-strand-break repair model for recombination. *Cell* **33**, 25-35.

Tabcharani J. A., Chang X., Riordan J. R., and Hanrahan J. W. (1991) Phosphorylation-regulated Cl- channel in CHO cells stably expressing the cystic fibrosis gene. *Nature* **352**, 628-631.

Tabcharani J. A., Rommens J. M., Hou Y.-X., Chang X.-B., Tsui L.-C., Riordan J. R., and Hanrahan J. W. (1993) Multi-ion pore behaviour in the CFTR chloride channel. *Nature* **366**, 79-82.

Tate P., Skarnes W., Nan X., and Bird A. (1994) Targeted mutation of a methylated DNA binding protein in the mouse. *Nineth International Workshop, Molecular Genetics of The Mouse.* **Abstract**.

The Cystic Fibrosis Genetic Analysis Consortium. (1990) Worldwide survey of the deltaF508 mutation- report from the cystic fibrosis genetic analysis consortium. *Am.J.Hum.Genet.* **47**, 354-359.

The Cystic Fibrosis Genetic Analysis Consortium (1994) Population variation of common cystic fibrosis mutations. *Hum.Mut.* **4**, 167-177.

Thomas K. R. and Capecchi M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512.

Thomas K. R., Folger K. R., and Capecchi M. R. (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419-428.

Thomas K. R., Deng C., and Capecchi M. R. (1992) High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. *Mol.Cell.Biol.* **12**, 2919-2923.

Thomas P. J., Shenbagamurthi P., Ysern X., and Pedersen P. L. (1991) Cystic fibrosis transmembrane conductance regulator: Nucleotide binding to a synthetic peptide. *Science* **251**, 555-557.

Thomas P. J., Shenbagamurthi P., Sondek J., Hullihen J. M., and Pedersen P. L. (1992) The cystic fibrosis transmembrane conductance regulator. Effects of the most common cystic fibrosis-causing mutation on the secondary structure and stability of a synthetic peptide. *J.Biol.Chem.* **267**, 5727-5730.

Tizzano E. F. and Buchwald M. (1992) Cystic fibrosis: beyond the gene to therapy. *J.Pediatr.* **120**, 337-349.

Tizzano E. F., Chitayat D., and Buchwald M. (1993) Cell-specific localization of CFTR mRNA shows developmentally regulated expression in human fetal tissues. *Hum.Mol.Gen.* **2**, 219-224.

Trapnell B. C., Chu C.-S., Paakko P. K., Banks T. C., Yoshimura K., Ferrans V. J., Chernick M. S., and Crystal R. G. (1991) Expression of the cystic fibrosis transmembrane conductance regulator gene in the respiratory tract of normal individuals and individuals with cystic fibrosis. *Proc.Natl.Acad.Sci.USA* **88**, 6565-6569.

Trezise A. E. O. and Buchwald M. (1991) In vivo cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature* **353**, 434-437.

Trezise A. E. O., Szirer C., and Buchwald M. (1992) Localization of the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) in the rat to chromosome 4 and implications for the evolution of mammalian chromosomes. *Genomics* **14**, 869-874.

Trezise A. E. O., Chambers J. A., Wardle J., Gould S., and Harris A. (1993a) Expression of the cystic fibrosis gene in human foetal tissues. *Hum.Mol.Gen.* **2**, 213-218.

Trezise A. E. O., Linder C. C., Grieger D., Thompson E. W., Meunier H., Griswold M. D., and Buchwald M. (1993b) CFTR expression is regulated during both the cycle of the seminiferous epithelium and the oestrous cycle of rodents. *Nat.Gen.* **3**, 157-164.

Tsui L.-C. (1992a) Mutations and sequence variations detected in the cystic fibrosis transmembrane conductance regulator (CFTR) gene:a report from the cystic fibrosis genetic analysis consortium. *Hum.Mut.* **1**, 197-203.

Tsui L.-C. (1992b) The spectrum of cystic fibrosis mutations. *TIG* **8**, 392-398.

Tsui L.-C., Buchwald M., and Barker D. (1985) Csytic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**, 1054-1057.

Tucker S. J., Tannahill D., and Higgins C. F. (1992) Identification and developmental expression of the Xenopus laevis cystic fibrosis transmembrane conductance regulator gene. *Hum.Mol.Gen.* **1**, 77-82.

Tyulewicz V. L. J., Tremblay M. L., LaMarca M. E., Willemsen R., Stubblefield B. K., Winfield S., Zablocka B., Sidransky E., Martin B. M., Huang S. P., Mintzer K. A., Westpal H., Mulligan R. C., and Ginns E. I. (1992) Animal model of Gaucher's disease from targeted disruption of the mouse glucocerebrosidase gene. *Nature* **357**, 407-410.

Valancius V. and Smithies O. (1991a) Double-strand gap repair in a mammalian gene targeting reaction. *Mol.Cell.Biol.* **11**, 4389-4397.

Valancius V. and Smithies O. (1991b) Testing an "in-out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Mol.Cell.Biol.* **11**, 1402-1408.

Valverde M. A., Diaz M., Sepulveda F. V., Gill D. R., Hyde S. C., and Higgins C. F. (1992) Volume-regulated chloride channels associated with the human multidrug-resistance P-glycoprotein. *Nature* **355**, 830-833.

Vieira T. and Messing J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259.

Wagner D. K. and Cavalli-Sforza L. L. (1975) Ethnic variation in genetic disease: possible roles of hitchhiking and epistasis. *Am.J.Hum.Genet.* **27**, 348-364.

Wainwright B. J., Scrambler P. J., and Schmidtke J. (1985) Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318**, 384-385.

Wainwright B. J., Scrambler P. J., and Scannier P. (1988) Isolation of a human gene with protein sequence similarity to human and murine int-1 and the Drosophila segment polarity and mutant wingless. *EMBO J.* **7**, 1743-1748.

Waldman A. S. (1994) The search for homology does not limit the rate of extrachromosomal homologous recombination in mammalian cells. *Genetics* **136**, 597-605.

Walker J. E., Saraste M., Runswick M. J., and Gay N. J. (1982) Distantly related sequences in the alpha-subunits and beta subunits of ATP synthase, Myosin, Kinases, and other ATP-requiring enzymes, and a common nucleotide binding fold. *EMBO J.* **1**, 945-951.

Ward C. L., Krouse M. E., Gruenert D. C., Kopito R. R., and Wine J. J. (1991) Cystic fibrosis gene expression is not correlated with rectifying Cl- channels. *Proc.Natl.Acad.Sci.USA* **88**, 5277-5281.

Watson E. K., Mayall E. S., Lamb J., Chapple J., and Williamson R. (1992) Psychological and social consequences of community carrier screening programme for cystic fibrosis. *Lancet* **340**, 217-220.

Welsh M. J. (1990) Abnormal regulation of ion channels in cytsic fibrosis epithelia. *FASEB.J.* **4**, 2718-2725.

Welsh M. J. and Liedtke C. M. (1986) Chloride and potassium channels in cystic fibrosis airway epithelia. *Nature* **322**, 467-470.

Welsh M. J. and Smith A. E. (1993) Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell* **73**, 1251-1254.

White R., Woodward S., and Zeppert M. (1985) A closely linked genetic marker for cystic fibrosis. *Nature* **318**, 382-384.

Whittingham D. G. (1971) Culture of mouse ova. *J. Reprod. Fertil.* **14**, 7-21.

Widdicombe J. H., Welsh M. J., and Finkbeiner W. E. (1985) Cystic fibrosis decreases the apical membrane chloride permeability of monolayers cultured from cells of tracheal epithelium. *Proc.Natl.Acad.Sci.USA* **82**, 6167-6171.

Williams R. L., Hilton D. J., Pease S., Wilson T. A., Stewart C. L., Gearing D. P., Wagner E. F., Metcalf D., Nicola N. A., and Gough N. M. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-687.

Windle J. J., Albert D. M., O'Brien J. M., Marcus D. M., Disteche C. M., Bernards R., and Mellon P. L. (1990) Retinoblastoma in transgenic mice. *Nature* **343**, 665-669.

Wine J. J. (1993) Indictment of pore behaviour. *Nature* **366**, 18-19.

Wine J. J. and Silverstein S. C. (1992) ATP and chloride conductance. *Nature* **360**, 18.

Wood S. A., Allen N. D., Rossant J., Auerbach A., and Nagy A. (1993) Non-injection methods for the production of embryonic stem cell-embryo chimaeras. *Nature* **365**, 87-89.

Wu C.-L. and Melton D. W. (1993) Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyltransferase-deficient mice. *Nat.Gen.* **3**, 235-239.

Wu H., Liu X., and Jaenisch R. (1994) Double replacement: strategy for efficient introduction of subtle mutations into the murine Col1a-1 gene by homologous recombination in embryonic stem cells. *Proc.Natl.Acad.Sci.USA* **91**, 2819-2823.

Wurst W., Auerbach A. B., and Joyner A. L. (1994) Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development* **120**, 2065-2075.

Yagi T., Aizawa S., Tokunaga T., Shigetani Y., Takeda N., and Ikawa Y. (1993) A role for fyn tyrosine kinase in the suckling behaviour of neonatal mice. *Nature* **366**, 742-745.

Yang Y., Devor D. C., Engelhardt J. F., Ernst F. A., Strong T. V., Collins F. S., Cohn J. A., Frizzell R. A., and Wilson J. M. (1993) Molecular basis of defective anion transport in L cells expressing recombinant forms of CFTR. *Hum.Mol.Gen.* **2**, 1253-1261.

Yorifuji T., Lemna W. K., Ballard C. F., Rosenbloom C. L., Rozmahel R., Plavsic N., Tsui L.-C., and Beaudet A. L. (1991) Molecular cloning and sequence analysis of the murine cDNA for the cystic fibrosis transmembrane conductance regulator. *Genomics* **10**, 547-550.

Zeitlin P. L., Crawford I., Lu L., Woel S., Cohen M. E., Donowitz M., Montrose M. H., Hamosh A., Cutting G. R., Gruenert D., Huganir R., Maloney P., and Guggino W. B. (1992) CFTR protein in primary and cultured epithelia. *Proc.Natl.Acad.Sci.USA* **89**, 344-347.

Zhang H., Hasty P., and Bradley A. (1994) Targeting frequency for deletion vectors in embryonic stem cells. *Mol.Cell.Biol.* **14**, 2404-2410.

Zheng H., Hasty P., Brenneman M. A., Grompe M., Gibbs R. A., Wilson J. H., and Bradley A. (1991) Fidelity of targeted recombination in human fibroblasts and murine embryonic stem cells. *Proc.Natl.Acad.Sci.USA* **88**, 8067-8071.

Zielenski J., Bozon D., Markiewicz D., Aubin G., Simard F., Rommens J. M., and Tsui L.-C. (1993a) Analysis of CFTR transcripts in nasal epithelial cells and lymphoblasts of a cystic fibrosis patient with 621+1G-T and 711+1G-T mutations. *Hum.Mol.Gen.* **2**, 683-687.

Zielenski J., Fujiwara T. M., Markiewicz D., Paradis A. J., Anacleto A. I., Richards B., Schwartz R. H., Klinger K. W., Tsui L. C., and Morgan K. (1993b) Identification of the M1101K mutation in the cystic fibrosis conductance regulator (CFTR) gene and complete detection of cystic fibrosis mutations in the Hutterite population. *Am.J.Hum.Genet.* **52**, 609-615.

Zimmer A. and Gruss P. (1989) Production of chimaeric mice containing embryonic stem (ES) cells carrying a homeobox Hox-1.1 allele mutated by homologous recombination. *Nature* **338**, 150-153.

APPENDIX 1

CYSTIC FIBROSIS

A1.1 CYSTIC FIBROSIS

This thesis describes an attempt to create mouse models for Cystic fibrosis by gene targeting. The following introduction seeks to give an overview of the nature of this disease, and provide some insight into the many aspects still to be resolved .

Cystic fibrosis (CF) is the most common fatal autosomal recessive genetic disease of the Caucasian population with about 1 in 2000 live births affected and therefore a carrier frequency of around 1 in 22. It is characterised by defective chloride ion transport and excessive mucus production in exocrine tissues which leads to a variety of clinical consequences (Boat *et al.* 1989).

CF was first definitely reported by Fanconi in 1936 and a detailed clinical description published by Andersen in 1938 describing the disease as 'cystic fibrosis of the pancreas'. In 1945 Farber recognised the role of mucus in CF and suggested the clinical symptoms were a consequence of disease of exocrine glands characterised by failure to clear mucus secretions. To reflect the mucus involvement the disease was renamed 'mucovisidosis', a name which is still in use in France, although it is has been known as cystic fibrosis in many countries since the 1960s. High salt concentration in saliva and sweat in CF was demonstrated by di Sant' Agnesi in 1953 and led to the development of the Gibson and Cooke 'sweat test', a diagnostic assay still in use today (Gibson and Cooke, 1959).

A1.1.1 Clinical and Pathological Features

The range and severity of symptoms associated with CF varies considerably and consequently so does the age of diagnosis. Previously, CF patients did not usually survive their teenage years and were cared for by paediatric specialists. Improvements in the care and management of CF have led to a concurrent improvement in life expectancy and it has been projected that a CF patient born in 1990 can expect an average life span of 40 years, double that of 20 years ago (Elborn *et al.* 1991).

Meconium ileus, a failure to pass meconium following birth, occurs in 5-10% of CF cases and is virtually diagnostic for the disease. It was thought to result from failure of pancreatic secretion leading to non-digestion of intraluminal contents also perhaps combined with dehydration, resulting in a bowel obstruction. However it has been shown that in some cases, the occurrence of meconium ileus does not always correlate with pancreatic insufficiency (Tizzano and Buchwald, 1992). Beyond the newborn period, bowel obstruction with incompletely digested contents can occur and is termed 'meconium ileus equivalent'. Treatment of meconium ileus usually involves a gastrografin enema but occasionally requires surgical intervention (Docherty *et al.* 1992).

In many cases of CF an absence of pancreatic secretions is observed as a consequence of obstruction of the pancreatic ducts with inspissated mucus followed by dilation and ultimately destruction of acinar cells and their replacement with fibrous tissue. The lack of pancreatic digestive enzymes leads to a variety of digestive disorders such as steatorrhoea which often manifests as a failure to gain weight. Diabetes, although endocrinic in nature, is sometimes a complication of CF and has been attributed to disturbance of islet architecture by acinar destruction and fibrosis. Pancreatic insufficiency in CF is rarely life-threatening and can be managed to a certain extent by pancreatic enzyme supplementation.

Lung infection and the subsequent tissue damage is the major cause of premature death in CF (Elborn and Shale 1990). There is great variation in age of onset, severity and course of lung disease between affected individuals. Although antibiotic treatment may suppress infection to a certain extent, it cannot easily be eradicated and there is an inevitable decline in lung function with time. Lung disease in CF arises as a result of the production of excessive viscous mucus which cannot be cleared from the lungs and so provides an environment which is easily colonised by opportunistic pathogens. Histological abnormalities such as submucosal gland hypertrophy and mucus cell hyperplasia are detectable within the first few days of life, prior to micro-organism colonisation, indicating that mucus accumulation is the

primary event. Repeated endobronchial infection initiates a continuous inflammatory response involving both humoral and cellular responses. Prolonged stimulation of host defences by micro-organisms and their antigenic by-products leads to the accumulation of cytotoxic by-products such as oxidants, proteases and lipid products, all of which are damaging to the pulmonary epithelia. Infection of this damaged tissue re-stimulates the inflammatory response leading to a vicious circle of infection, inflammation, and lung injury. Chronic pulmonary infection and lung injury eventually culminates in respiratory failure associated with pulmonary hypertension and cor pulmonale.

Fertility is much reduced in CF patients, especially in the male. Greater than 95% of males have altered Wolffian duct structures, commonly congenital bilateral absence of the vas deferens (CBAVD) and are therefore azoospermic (Denning *et al.* 1968., Kaplan *et al.* 1968). It has been suggested that this might arise as a result of blockage of the Wolffian duct structures with inspissated secretions during foetal development causing gradual occlusion of the male genital tract. Motile sperm are however present in the epididymis of patients with this condition, and fertilisation can be achieved *in vitro* through use of the Microscopic Epididymal Sperm Aspiration (MESA) technique (Oates *et al.* 1992). Congenital bilateral absence of the vas deferens accounts for 1-2% of all cases of male infertility (Jequier *et al.* 1985), and it has been reported that a high number (64%) of healthy but infertile males presenting with CBAVD at infertility clinics carried at least one detectable CF mutation (Anguiano *et al.* 1992., Rigot *et al.* 1991). This has lead to the suggestion that these cases represented a mild, primarily genital form of CF, and the recommendation that CF mutational analysis should be carried out on any patients with CBAVD and their families, prior to sperm aspiration to remedy infertility. The relationship between CF and CBAVD has been further investigated by Osborne *et al.* (1993), who also found an increased frequency of *CFTR* mutations in CBAVD patients, with the majority having only one detectable CF mutation. The bioelectric properties of the nasal epithelium of these patients indicated a defect in chloride ion transport, however this defect was distinct from that displayed by individuals with CF. They suggest that

these patients might harbour a different class of mutations such as those in the promoter region or at other regulatory sites which affect the expression level of the gene and are not so easily detected. The low level of *CFTR* expression resulting from such mutations might be sufficient to prevent disease in all but the most susceptible organs such as the testis, where *CFTR* is expressed at all stages of human foetal development (Tizzano *et al.* 1993., Trezise *et al.* 1993). Therefore the relationship between CBAVD and CF is still unclear, but CFTR remains implicated in its aetiology.

Female CF patients are generally more fertile than males with around 75% able to sustain pregnancies (Cohen *et al.* 1980). Poor health status and thick cervical mucus often hinder conception, and the physical stress imposed by pregnancy is often detrimental to the health of the mother (Boat 1989., Kotloff *et al.* 1992).

Other complications of CF can include cirrhosis of the liver and arthropathy (a form of joint disease). The psychosocial consequences of living with a life-threatening and incurable disease which imposes a demanding daily regime of physiotherapy and drug therapy are far reaching. The poor prognosis of patients contracting the highly infective *Pseudomonas cepacia* (recently renamed *Burkholdaria cepacia*), has led to the segregation of patients with positive sputum cultures both inside and outside of the hospital environment (Smith *et al.* 1993., Govan *et al.* 1993) This is particularly stressful as previously, patients had been encouraged to socialise and many facilities and associations have been set up to encourage this. Living with infertility can also have psychological complications, especially if the patient is in relatively good health.

Diagnosis of CF is usually suggested from the clinical features and then confirmed by the sweat test (Gibson and Cooke 1959), and/or a test for immunoreactive trypsin (IRT) (Ryley *et al.* 1981). For the sweat test, sweating is induced by passing a small current into the skin of the forearm through a pilocarpine-soaked filter paper. The resulting sweat is collected, 100 mg of which is needed for a reliable result, and the

chloride concentration measured. A concentration greater than 60 meq/litre is indicative of CF but not in itself diagnostic. The IRT test was developed primarily for screening neonates where difficulty in obtaining sufficient quantities of sweat is often experienced. The presence of immunoreactive trypsin in the blood caused by leak-back of trypsin from an obstructed pancreas is detected using immunoassay techniques.

A1.1.2 The Basic Defect In CF

The first insight into the basic biochemical defect of CF arose in 1981 when Knowles *et al.* reported that the potential differences (PD) across nasal and airway epithelia in CF patients of all ages and stages of the disease were elevated, approximately twice that of normal controls. The application of amiloride, an inhibitor of active sodium absorption, resulted in a decrease in PD in both normal and CF epithelia. However this decrease was much greater in CF epithelia, indicating that sodium absorption was elevated in CF. The role of chloride transport in CF became apparent in 1983 when Quinton demonstrated abnormally low chloride permeability in CF sweat ducts. This would result in poor sodium chloride reabsorption from sweat and thus accounts for the high sweat salt concentration characteristic of CF. This result was quickly followed by demonstration of reduced chloride permeability in CF nasal mucosa (Knowles *et al.* 1983).

Since these early studies, reduced chloride permeability has been shown in most affected epithelia (Welsh 1990), and increased sodium transport in airway epithelia (Knowles *et al.* 1986., Cotton *et al.* 1987). The defective chloride permeability was localised to the apical membrane (Widdicombe *et al.* 1985), and results *in vivo* confirmed by experimentation with primary and immortalised cultures of CF epithelial cells *in vitro* (Stutts *et al.* 1985., Jetten *et al.* 1989).

In 1986, Welsh and Liedtke suggested that the chloride impermeability was a result of defective chloride channel regulation as they had found that normal and CF airway epithelia contain chloride channels with identical properties, however, chloride

channels in CF epithelia (unlike normal epithelia) did not open in patch-clamp experiments in the cell attached form. Furthermore, the isoprenaline-stimulated increase in intracellular cAMP response was normal in both, and the calcium-dependent stimulus-response coupling was intact as demonstrated by activation of calcium-dependent potassium channels. Thus they concluded that regulation of chloride channels at a site distal to cAMP accumulation is defective in CF.

Using the Patch-Clamp technique, Frizzell *et al.* (1986), showed that the β-adrenergic-stimulated cAMP-induced activation of chloride channel activity was defective in CF, and suggested that defective phosphorylation of the channel by a protein kinase may be involved. The role of protein kinase phosphorylation in chloride channel activation became apparent when it was shown that the catalytic subunit of cAMP-dependent protein kinase plus ATP could activate chloride channels in excised membrane patches of normal, but not CF epithelia (Schoumacher *et al.* 1987., Li *et al.* 1988). This suggested that the block in cAMP-mediated activation of CF chloride channels lay distal to induction of cAMP-dependent protein kinase activation, and so might be a defect in the chloride channel itself or an associated regulatory protein. Further work using patch-clamp (Hwang *et al.* 1989) or $^{125}I^-$ efflux techniques (Li *et al.* 1989) to measure channel activation, showed that both protein kinase A and protein kinase C could activate chloride channels in normal airway epithelia but not CF airway epithelia.

In September 1989, the study of CF changed dramatically with the cloning of the CF gene (Rommens *et al.* 1989., Riordan *et al.* 1989., Kerem, *et al.* 1989). Initially, homology to a family of transport proteins suggested that the gene product (*CFTR*) was itself a transporter and not a chloride channel (Hyde *et al.* 1990., Ringe and Petsko, 1990). The subsequent availability of *CFTR* cDNA spawned a whole host of expression studies involving the expression of *CFTR* in many different cell types *in vitro* which suggested that CFTR was in fact a chloride channel (Rich *et al.* 1990., Kartner *et al.* 1991., Bear *et al.* 1991., Anderson *et al.* 1991., Tabcharani *et al.*

1991), as expression of *CFTR* in these cells coincided with the appearance of regulated chloride channel activity not previously detected.

This classification of CFTR as a chloride ion channel created more controversy. An ion channel has certain unique features including ion selectivity, single channel conductance, and blocker sensitivity, by which it can be identified, however, previous studies had not all agreed on the 'signature' of the CFTR chloride channel. Some studies reported a 30pS-50pS outwardly-rectifying chloride channel (ORCC), and others a much smaller 10pS chloride channel with a linear current-voltage relationship (reviewed in Frizzell and Cliff 1991). Three different chloride conductances had been identified in chloride-secreting epithelial cells (Cliff and Frizzell, 1990), and it was unclear which was defective in CF (the possibility of CFTR being responsible for more than one conductance deemed unlikely).

Doubt was cast on the role of the outwardly-rectifying chloride channel in CF, when it was shown that there was no correlation between *CFTR* expression and the density of ORCC in human epithelial cell lines (Ward *et al.* 1991). Further proof that CFTR is a chloride channel was demonstrated by the appearance of new chloride channel activity in many different cell types *in vitro* when transfected with *CFTR* cDNA, also brought forward evidence against the ORCC (Tabcharani *et al.* 1991., Rich *et al.* 1991., Anderson *et al.* 1991., Bear *et al.* 1991). Characterisation of the single channel properties of chloride channels in cells expressing recombinant *CFTR* by Berger *et al.* (1991), identified a 10pS chloride channel with a linear current-voltage relationship.

Definitive proof that CFTR was indeed a chloride channel came from the reconstitution studies of Bear *et al.* (1992), who studied the function of highly purified CFTR when incorporated into an artificial liposome bilayer. The chloride channel activity thus demonstrated ended not only speculation on the function of CFTR, but also the role of ORCC in CF as the chloride channel exhibited the 10pS current and the linear current-voltage relationship of the non-rectifying channel.

Therefore it was concluded that the ORCC probably had little to do with the CFTR chloride channel, and that earlier observations were 'unfortunate' (Higgins and Hyde, 1991).

This state of affairs remained until Egan *et al.* (1992) published evidence that not only are CFTR and the ORCC distinct chloride channels, but CFTR can regulate the ORCC. They arrived at this conclusion from the observation that the ORCC could not be activated by protein kinase A in cells lacking CFTR, but this activation was restored on expression of *CFTR*. Subsequent studies on CF 'knockout' mice which lack *CFTR* (discussed later), reported the presence of ORCC in mouse epithelial cells indicating that CFTR and ORCC are indeed distinct chloride channels (Gabriel *et al.* 1993). The failure of protein kinase A to regulate ORCC in these epithelial cells lacking CFTR further supports the suggestion that they are regulated by CFTR. The early data of Welsh and Liedtke (1986) and Frizzell (1986) which suggested that it is regulation of the CF chloride channel and not the channel itself which is defective, can perhaps be explained by them studying the ORCC instead of CFTR.

Therefore, the basis of CF is defective chloride ion transport by the 10pS chloride ion channel which has a linear current-voltage relationship. A defect in this chloride ion channel may have many indirect but far-reaching effects including defective regulation of other chloride ion channels.

A1.1.3 Cloning The CF Gene

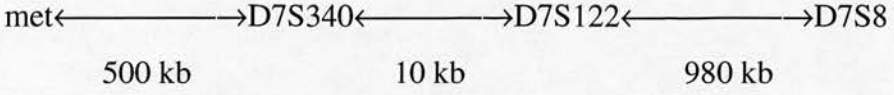
The CF gene was successfully cloned by a method known as 'positional cloning' or 'reverse genetics', which searches for linkage of a chromosomal region with the disease in question, and does not require a knowledge of the biochemical nature of the disease.

The first steps towards assignment of the CF gene to a particular chromosome arose when linkage was demonstrated to the enzyme paraoxonase (Eiberg *et al.* 1985), followed by linkage to the polymorphic marker DOCR1-917 (Tsui *et al.* 1985).

Further polymorphic markers, met and pJ3.11 (D7S8) were isolated, and the CF gene located to the long arm of chromosome 7 (White *et al.* 1985., Wainwright *et al.* 1985). A collaboration of several groups typing over 200 families showed that the *met* and D7S8 markers flanked the CF gene (Beaudet *et al.* 1986). This set the stage for novel cloning strategies to begin searching between these markers which were separated by 1600 kb of DNA.

One such attempt used *met* oncogenic transformation of cells in chromosome-mediated gene transfer to produce a somatic cell hybrid containing 3 Mbp of human DNA from the supposed CF locus (Scambler *et al.* 1986). This was then scanned for the presence of CpG islands which are characteristic of the 5' end of genes (Bird 1986). With this method they found a candidate gene at the predicted location which they named the *Int-1* related protein (Estivill *et al.* 1987). However this gene did not meet the criteria for validation of a candidate gene as the nature of the gene product, a secreted growth factor, and it's pattern of expression did not fit that predicted for the CF gene, and furthermore no mutations in this gene in CF patients could be detected (Wainwright *et al.* 1988). Then conclusive proof that this was not the CF gene arose when recombination between the gene and CF was reported (Berger *et al.* 1987, Farrall *et al.* 1988).

Using saturation cloning, Rommens *et al.* (1988), discovered two more polymorphic markers which lay inside the *met* and D7S8 markers and determined that the orders and distances were as followed:



Using these markers as the starting points, Collins *et al.* (1987) used the chromosome jumping technique to move between the markers. The first two jumps brought them to the CpG island of the *Int-1* gene, confirmation that they were moving in the right direction. Further jumps 'filled in' with chromosome walking led them to a region

which had maximum linkage disequilibrium with CF. Within this region a search was conducted for sequences which showed cross-species conservation indicative of functional DNA. This produced a candidate gene segment encoding a 6.2 kb transcript which had a short sequence of open reading frame preceded by a CpG island, and therefore some hallmarks of a gene (Rommens *et al.* 1989). Despite these favourable signs, the RNA transcript could not be detected in tissues predicted to express the CF gene product. The breakthrough came when a sweat gland cDNA library was searched and resulted in the identification of a 6.5 kb transcript with a pattern of expression that was restricted to exocrine tissues. Sequentially isolated cDNAs eventually led to a series of overlapping clones containing the entire coding region of the gene but construction of a complete cDNA was very problematic (Riordan *et al.* 1989). This gene was encoded by 250 kb of DNA, with 27 exons, and from which a protein of 1480 amino acids was translated. Expression was found to be higher in those tissues severely affected in CF, but no size difference was detected in the CF transcript. Confirmation that this was indeed the CF gene was demonstrated by the presence of mutations in this gene of individuals with CF but not in normal controls. Therefore this gene was validated as the CF gene as it fulfilled three essential criteria;

1. The gene was in the correct location and had no detectable recombination with the disease.
2. The nature of the gene product was consistent with the disease and had an appropriate pattern of expression.
3. Consistent mutations were found within this gene. The disequilibrium for one particular haplotype had suggested a common mutation, and the Δ F508 mutation was subsequently found to be present on 70% of CF chromosomes.

A1.1.4 The Structure Of CFTR

Sequence analysis of the CF gene, named the Cystic Fibrosis Transmembrane Conductance Regulator, (*CFTR*) predicted an encoded protein of 1480 amino acids

with a molecular mass of 168 kd. The main structural features of this protein (Figure 1.1) deduced from amino acid sequence analysis were characteristic of a membrane bound protein due to the presence of two hydrophobic repeated motifs consisting of six α-helical sequences, typical of membrane spanning domains (MSD) (Riordan *et al.* 1989). Each membrane spanning domain is followed by a large hydrophilic region predicted to lie on the cytoplasmic side of the membrane, which has sequences resembling nucleotide binding domains (NBDs). These structural characteristics of CFTR have similarities to a family of ATP-dependent transport proteins called the ATP binding cassette proteins (ABC). The most notable similarity being to the multiple drug resistant (MDR) proteins and the yeast STE6 protein, all of which import or export molecules across cell membranes in a process requiring ATP hydrolysis (Higgins 1989). There is one feature of CFTR which appears to be unique, that of the presence of a highly charged hydrophilic region lying in the middle of the protein in the cytoplasmic domain, the R domain. This region contains many polar residues and has multiple potential phosphorylation sites.

A1.1.5 The Function Of CFTR

CFTR is a chloride channel. This is the conclusion that has been reached over the period since the CF gene was cloned. There has been much debate over the function of CFTR, and even now that there is general agreement on the role of CFTR as a chloride channel, the prospect of additional functions for CFTR has arisen.

The first evidence supporting the role of CFTR as a chloride channel was put forward by Rich *et al.* 1990, who showed that expression of *CFTR*, but not mutant *CFTR*, corrects the chloride transport defect in cultured CF epithelial cells. Expression of *CFTR in vitro* was shown to generate a cAMP activated chloride channel which was absent in those transfected with ΔF508 cDNA and in nontransfected controls (Anderson *et al.* 1991). Similarly, Kartner *et al.* (1991) reported generation of a chloride channel in Sf9 insect cells upon transfection with *CFTR* cDNA, and Tabcharani *et al.* (1991) observed the appearance of a chloride channel activity in CHO cells upon stable transfection with *CFTR* cDNA. In addition, regulation of the

CHO channel was shown to be exerted by cAMP-dependent phosphorylation by protein kinases. Translation of CFTR RNA in the *Xenopus* oocyte expression system resulted in a chloride selective current arising from a channel activated by protein kinase A (Bear *et al.* 1991). Translation of this RNA had no affect on the activity of the endogenous chloride channels and so strengthened the argument that CFTR itself was a chloride channel, rather than a regulatory protein which modulates other chloride channels.

The anion selectivity of CFTR-generated chloride channels was determined by Anderson *et al.* (1991). They found that the introduction of mutations into transmembrane domain-coding sequences altered the anion selectivity of the chloride channel, and was therefore consistent with the role of CFTR as a chloride channel.

The most convincing evidence was that of Bear *et al.* (1992), who reconstituted highly purified CFTR into artificial lipid bilayers. Using patch-clamp analyses, they observed cAMP-regulated currents similar to those recorded for CFTR-transfected cells, which could only be attributed to CFTR as no other protein was present.

If CFTR is a chloride channel, how does it carry out this function? It is thought that the transmembrane domains form a pore across the plasma membrane through which ions can pass. Alteration of the amino acids within this domain alters the anion selectivity of this channel (Anderson *et al.* 1992), indicating that the transmembrane domain both selects and conducts ions across the membrane.

The R domain, which lies between the two membrane spanning domains, is a feature of CFTR which is not shared by any known ABC protein. It is thought to regulate the chloride channel activity of CFTR, as its deletion generates chloride channels which are constitutively open (Rich *et al.* 1991). The poor cross species conservation of this region (Tucker *et al.* 1992), and the observation that no missense mutations altering the R domain have been associated with CF, suggests that the mechanism of channel activity regulation by this region is non-specific. It has been proposed that

CFTR channel activity is inhibited by a conformational change in the R domain, resulting in 'plugging' of the ion pore (Rich *et al.* 1991., Kartner *et al.* 1991., Cheng *et al.* 1991). This has been supported by the recent work of Dulhanty and Riordan (1994), who observed a structural change in CFTR following phosphorylation, which was indicative of a reduction in the α-helical content. The R domain is highly charged, and potential phosphorylation sites have been identified which if mutated, prevent cAMP dependent phosphorylation and activation of the chloride channel (Cheng *et al.* 1991). However, phosphorylation of the R domain alone does not activate the chloride channel, as binding of ATP by the nucleotide binding domains also appears to be required (Anderson *et al.* 1991). Furthermore, not all of the phosphorylation sites are required for activation of channel activity, and mutational studies have shown that the presence of a single site is sufficient for channel activation (Cheng *et al.* 1991).

The nucleotide binding domains (NBDs) which follow each membrane spanning domain, share extensive homology with the NBDs of other ABC transporter proteins (Higgins 1989). They contain Walker A and B sequences which interact with nucleotides (Walker *et al.* 1982), and have been shown to interact with ATP to regulate the opening and closing of CFTR chloride channels (Thomas *et al.* 1991., Anderson *et al.* 1991., Hartman *et al.* 1992). It has been found that there is a concentration of missense mutations in the first NBD, and a lack of a similar cluster in the second NBD, suggesting a difference in functional importance (Tsui and Buchwald 1991., Kerem *et al.* 1990., Cutting *et al.* 1990). Missense mutations are particularly informative as only one amino acid is altered in the mutant protein, and the observation that almost all missense mutations in *CFTR* occur in regions which are conserved between human, mouse and cow emphasise their functional significance (Dean *et al.* 1990., Diamond *et al.* 1991). This is supported by the expression studies of Gregory *et al.* (1991), studying the effect of CF mutations on the maturation and function of CFTR in Cos cells. They found that missense mutations in the first NBD almost always resulted in a lack of maturation or functional activity, but there was rarely any adverse effect seen with mutations in the

second NBD. Anderson and Welsh (1992) described further evidence for the functional non-equivalency of NBD1 and NBD2. They demonstrated that ATP can bind to both NBDs, but that ADP binds to NBD2 only. This resulted in inhibition of chloride channel activity, which was abolished by mutations in NBD2. They suggested that ADP inhibits CFTR by competing with ATP at NBD2, and that the ATP : ADP ratio may be more important than absolute ATP concentration for regulating CFTR. Quinton and Reddy (1992) developed this hypothesis further by demonstrating that nonhydrolysable ATP analogues could stimulate CFTR in human sweat ducts following activation by cAMP, suggesting that binding rather than hydrolysis of ATP was required for CFTR chloride channel activity. Coupled with the observation that intracellular ATP concentrations must be around physiological concentrations for activation by this apparent non-hydolysable mechanism, this has led to the idea that by binding ATP and ADP, CFTR is either 'sensing' the energy level of the cell (ATP Vs ADP), or the cellular ATP concentration itself and it is this which regulates CFTR activity (Wine and Silverstein 1992).

The significance of the repeated structural motifs of CFTR was investigated by Sheppard *et al.* (1994). They investigated the functionality of mutant CFTR consisting of MSD1 and NBD1, plus the R domain as some members of the ABC transporter protein family are known to contain a single membrane spanning domain and a single nucleotide binding domain (Hyde *et al.* 1990). They found that this mutant CFTR was capable of forming a regulated chloride channel which had many properties in common with wildtype CFTR. The mutant channel, like wild type, was regulated by PKA phosphorylation, although some activity was seen in the absence of PKA phosphorylation. This activation was reminiscent of that seen with mutant CFTR lacking part of the R domain (Rich *et al.* 1991), and suggests that the R domain interacts with NBD2 which was absent in this mutant form of CFTR. An inhibitory role for NBD2 was implied by the observation that ATP was more potent in stimulating channel activity in the mutant protein lacking NBD2. The ion conducting properties of this mutant CFTR were similar to those of wildtype CFTR, exhibiting identical anion selectivity, anion-to-cation permeability, single channel

conductance, and linear current-voltage relationship. This implies that the sequences in MSD1 are sufficient to form a regulated chloride ion channel, and questions the function of NBD2. Sheppard *et al.* (1994) speculate that NBD2 may be important for stabilising CFTR or localising it to the apical membrane, due to the observation that in this expression system fewer channels were generated by the mutant CFTR than compared to wildtype, despite production of large amounts of the mutant protein. However, these results suggest that the majority of sequences required to form a regulated chloride ion channel are present in the amino-terminal portion of CFTR.

Therefore it does appear that CFTR is a chloride channel, but does it have additional functions and are chloride ions the only molecules transported? The closely related human multidrug resistance P-glycoprotein has many structural similarities with CFTR, and belongs to the same group of ATP binding cassette (ABC) proteins. This protein has been found to be bifunctional, acting both as a pump which actively transports cytotoxic drugs out of the cell, and as a volume-regulated chloride channel (Valverde *et al.* 1992., Gill *et al.* 1992). By analogy, this raises the possibility of a dual function for CFTR. Hasegawa *et al.* (1992), demonstrated that CFTR can form a multifunctional aqueous channel capable of transporting anions, water and small solutes, and this was followed by evidence that fluid transport across airway epithelia was indeed defective in CF (Jiang *et al.* 1993). Tabcharani *et al.* (1993), have shown that CFTR functions as a multi-ion pore, transporting more than one anion simultaneously. In addition, they demonstrated that some naturally occurring CF mutations associated with a mild phenotype, reduces the channel to a single ion pore with a concurrent 50% reduction in conductance. Therefore it appears that CFTR may have more roles than simply transporting chloride ions.

CFTR has also been implicated in other cellular functions. Bradbury *et al.* (1992) observed an absence of cAMP-dependent regulation of endocytosis and exocytosis in pancreatic cells derived from a CF patient, which was restored upon transfection with wildtype *CFTR* cDNA. This suggests a role for CFTR in regulation of membrane recycling which is an important way for cells to control both secretion and placement

of proteins on their surface (Baringa, 1992). Barasch *et al.* (1991), reported that acidification of intracellular organelles such as the Golgi network, prelysosomes and endosomes was also defective in CF cells, and Lukacs *et al.* (1992) have reported that CFTR is present and functional in endosomes. The possibility that CFTR may be multifunctional may be important for understanding the relationship between this single gene defect and the many facets of this disease, as many of the clinical manifestations cannot be directly attributed to a simple defect in chloride ion transport. In particular, the abnormal composition of CF mucus which lacks sialic acid in the oligosaccharides, and the increased sulphation does not correlate directly with abnormal chloride ion transport (Baringa 1992., Richardson and Alton 1993), and might be more easily attributed to defective acidification of intracellular organelles.

A1.1.6 CF Mutations

A1.1.6.1 The spectrum of CF mutations.

Following identification of the CF gene, a consortium was formed, The Cystic Fibrosis Genetic Analysis Consortium, consisting of 90 laboratories from 26 countries. This was set up to coordinate the accumulation of information on the identification of new mutations and their population frequencies resulting in a large pool of information. To date, over 400 disease mutations have been identified (, The Cystic Fibrosis Genetic Analysis Consortium 1994), almost half of which are missense mutations. Non-sense, frameshift, and splice site mutations occur in roughly equal proportions, but few large deletions have been reported. No promoter mutations have been found, although this probably reflects a bias of mutation detection methods rather than an absence of such mutations.

Of all mutant CF chromosomes examined so far, 67% have been found to carry a three base pair (phenylalanine) deletion at position 508 (ΔF508). The frequency of this mutation varies in each population, and there is a clear Northwest to Southeast gradient of its frequency across Europe (De Braekeleer and Daigneault 1992), with the highest frequency reported in Copenhagen and Stockholm (CF Genetic Analysis

Consortium 1990). A study of ΔF508 chromosomes from fifteen European regions using three intragenic markers demonstrated that this mutation has a single origin, and probably arose more than 52,000 years ago (Morral *et al.* 1994). Haplotype analysis suggests that the ΔF508 mutation originated in a population genetically distinct from any present day European population, and spread into Europe at different periods, hence the differing frequencies in Europe. It has been proposed that the mutation was first introduced during the Palaeolithic period, which may account for the high frequency of this mutation (80% CF chromosomes) associated with the ancestral haplotype present in Basques which are thought to be a Palaeolithic population (Bertranpetit and Cavalli-Sforza 1991., Casals *et al.* 1992). The presence of the ΔF508 mutation at the high frequency of 1.4% (2% of chromosomes carry a CF mutation, 70% of which are the ΔF508 mutation) might be attributed to a founder effect, or alternatively, may imply an advantage to being a carrier, termed 'heterozygote advantage' (Knudson *et al.* 1967). This could be ascribed directly to the presence of a single CF mutation, or to an advantage conferred by closely linked loci (Wagner and Cavalli-Sforza 1975). It has been proposed that CF heterozygotes may have an advantage in surviving Cholera which is characterised by excessive secretion of chloride and fluid in the intestine in response to the Cholera toxin (Rodman and Zamudio 1991). This was tested by Gabriel *et al.* (1994) using a mouse model for CF. They observed that mice homozygous for a null mutation in *CFTR* did not secrete fluid in response to Cholera toxin, but mice heterozygous for the null mutation secreted only 50% of the fluid volume lost by wildtype mice in response to the toxin. This reduced response in heterozygotes implies that they may be better able to survive the diarrhoea and resultant water loss associated with Cholera.

Different mutations can predominate in different populations, e.g. 90% of CF mutations in the Danish population are ΔF508 (CF Genetic Analysis Consortium 1990), whereas only 22% are accounted for by ΔF508 in the Ashkenazi Jewish population where the W1282X mutation predominates (approximately 60% of CF chromosomes), (Shoshani *et al.* 1992., Lerer *et al.* 1992). The isolated nature of

some communities contributes to the high frequency of certain CF mutations within these communities which is not seen in the general population, as demonstrated by the Hutterite community where the otherwise extremely rare M1101K mutation accounts for 69% of CF chromosomes (Zielenski *et al.* 1993).

A1.1.6.2 Screening for CF mutations.

The large number of CF mutations and their variant frequencies makes detection of 100% of mutations within a population impossible, and therefore creates a dilemma for advocates of population screening programmes. Initially screening could detect only 70% of CF carriers thus identifying only 50% of at risk couples, a level that many found unacceptably low (Roberts 1990., Beaudet 1990., Gilbert 1989). Significant improvements have been made in the detection of CF mutations in various populations (Shoshani *et al.* 1992., Super and Schwarz 1992., Cheadle *et al.* 1992), with the highest rate of detection of 98% reported by Ferec *et al.* (1992) in a Celtic population. Pilot studies have reported a favourable response to CF carrier screening (Watson 1991., Decruyenaere *et al.* 1992) although this varied with cultural beliefs (Miller *et al.* 1992). There is great debate over how screening should be delivered to the population. Screening the population in general has the advantage that carriers are identified before pregnancy and so have time to consider the consequences of being a CF carrier and decide their future reproductive plans. However, with time, the significance of being a carrier may be forgotten or confused, and this type of screening often has a low take-up rate (Watson *et al.* 1991). Another mode, that of screening women attending antenatal clinics have been found to have a much higher take-up rate (Mennie *et al.* 1992), but has the disadvantage that carriers are only identified after a pregnancy has been initiated, and allows very little time for decision-making. A third type of screening, commonly called 'cascade' screening, uses a CF patient as a starting point from which all immediate relatives are tested. The relatives of anyone testing positive are subsequently tested, and so on, resulting in a cascade effect (Super *et al.* 1992). The advantage of this method is the high take-up rate, and the targeting of screening to those most at risk. Whatever the method of screening, the importance of effective counselling is shown in a study of

Marteau *et al.* (1992) who found that carriers of the recessive genetic disease Tay-Sachs identified in a screening programme viewed their future health prospects in more negative terms. Also, a pilot screening project in Greece for Haemoglobinopathies (Stamatoyannopoulus 1974) resulted in stigmatisation of carriers by their community and they were considered only suitable for marriage by other carriers!

A1.1.6.3 Correlation of genotype with phenotype.

There is a great range of symptoms and degrees of severity associated with CF, and their correlation with a particular genotype has often been attempted. The ability to predict the course and severity of the disease for each individual would be of enormous value both for informed decision-making following a positive prenatal diagnosis for CF, and for counselling and care of an affected individual. A recent report from The Cystic Fibrosis Genotype-Phenotype Consortium (1993), attempted to correlate the phenotypes of approximately 62% of CF patients with their genotype. They concluded that 'the only prognostic value of genotypic information is for the prediction of pancreatic status'. The association that has become apparent is the correlation of so called 'severe' alleles and pancreatic insufficiency (PI) (Kristidis *et al.* 1992). One such severe allele is ΔF508 which, in the homozygous state, has been frequently correlated with pancreatic insufficiency (Kerem *et al.* 1990., Santis *et al.* 1990., Pignatti 1991). When CF individuals are homozygous for a severe mutation or have two different severe mutations, then they are usually PI. Individuals with a severe and a 'mild' mutation e.g. R117H, are usually pancreatic sufficient (Johansen *et al.* 1991., Kristidis *et al.* 1992., Osbourne *et al.* 1992., Ferec *et al.* 1993). The mild mutation being dominant over the severe mutation. Four mutations have been designated severe, these are ΔF508, G542X, R553X, and W1282X, and prenatal and prognostic counselling for homozygotes and compound heterozygotes should advise of longterm PI (CF Genotype-Phenotype Consortium, 1993). A reduced risk of meconium ileus has been reported for the G551D mutation (Hamosh *et al.* 1992). Homozygosity for ΔF508 has also been associated with an earlier age of diagnosis (Kerem *et al.* 1990., Hamosh *et al.* 1992), but has not been found to correlate with

severity of pulmonary disease (Santamaria *et al.* 1992., Santis *et al.* 1990). The conclusion of the CF Genotype-Phenotype Consortium is that severity and course of pulmonary disease are not predicted by genotype, and they suggest that factors other than CF genotype affect the pulmonary phenotype. Kiesewetter *et al.* (1993) reported that the variation of the phenotype of CF patients carrying at least one R117H mutation in association with another CF mutation correlated with the chromosomal background. Correlation between CF mutations and other aspects of CF have been attempted but no clear picture has emerged. Dodge (1991) cautions that the contribution of non-genetic factors must not be overlooked, especially as identical twins with different health status' have been reported, and that there will always be great variability in exposure to infection.

A1.1.6.4 The effect of CF mutations on CFTR function.

The consequences of ΔF508 mutation on *CFTR* function has been extensively studied. Cheng *et al.* (1990) studied the expression of mutant ΔF508 *CFTR* cDNA in Cos cells and observed that mature CFTR was absent in these cells. Instead, an incompletely glycosylated form was detected which they suggested was a result of incomplete CFTR processing in the endoplasmic reticulum (ER) resulting in its degradation and therefore absence at the correct cellular location. The ΔF508 allele has been shown to be expressed at the same level as the normal allele in the respiratory tract (Trapnell *et al.* 1991) indicating that transcription of mutant *CFTR* is not defective. Further study into the processing and intracellular location of ΔF508 CFTR produced conflicting results. Studies examining the maturation and function of mutant *CFTR* expressed in transfected cells *in vitro* (Gregory *et al.* 1991., Dalemans *et al.* 1992), agreed with those of Cheng *et al.* Nevertheless, other studies detected correct localisation of ΔF508 CFTR in the plasma membrane (Sarkadi *et al.* 1992), and chloride channel activity when ΔF508 cDNA was expressed in *Xenopus* oocytes (Drumm *et al.* 1991), Vero cells (Dalemans *et al.* 1991), and Sf9 insect cells (Bear *et al.* 1992). A subsequent publication was able to reconcile these differences by demonstrating that the processing of mutant CFTR is temperature sensitive (Denning *et al.* 1992). They showed that processing of ΔF508 CFTR reverts to wild

type as the temperature is lowered, thereby explaining the differences seen in processing and localisation in experiments using oocytes and insect cells which are typically conducted at lower temperatures. The anomalous results of the experiments using Vero cells were obtained in a very high expression system and therefore might be a result of some protein escaping ER 'quality control mechanisms' and reaching the plasma membrane where it could generate some chloride channel activity. This was confirmed by an important observation that the ΔF508 mutation was also mislocalised in sweat gland biopsies taken from CF patients (Kartner *et al.* 1992). This led to speculation that the basis of defective CFTR channel activity of ΔF508 mutants might be entirely attributable to its mislocalisation which if overcome, might have sufficient activity to correct the defect. This was confirmed by Li *et al.* (1993) who demonstrated that ΔF508 mutant CFTR exhibited phosphorylation-regulated chloride channel activity similar to that of wildtype CFTR when reconstituted into lipid bilayers. Yang *et al.* (1993), followed this by showing that incubation of L cells transfected with ΔF508 cDNA at 26°C for 48 hours elicited an electrophysiological response which was absent in cells incubated at the higher temperature of 37°C. This observation is of great importance for the treatment of CF, as it implies that if ΔF508 CFTR could be relocated to the plasma membrane, the mutant protein should exhibit enough chloride channel activity to ameliorate or even correct the CF defect.

Although defective processing is also observed with other CF mutations, this is not the sole mechanism by which CF-causing mutations exert an effect. Welsh and Smith (1993) have divided CF mutations into four classes based upon their effect on CFTR biosynthesis and function (table 1.1, and figure 1.2):

Class I includes all CF mutations which result in defective protein production. This might be due to the presence of premature termination signals, the production of unstable mRNA, or severely truncated or aberrant protein which is unstable and quickly degraded. All mutations of this type have the net result that no detectable CFTR protein is produced.

Class II mutations are those which fail to progress through the biosynthetic pathway and traffic to the correct cellular location, and includes the ΔF508 mutation. The basis of the defect caused by this class of mutations is the absence of CFTR at the plasma membrane.

Class III mutations are processed correctly, and do reach the plasma membrane, but exhibit defective regulation of the chloride channel activity. These mutations have been found to be located in the nucleotide binding domains, and NBD1 in particular which is consistent with its proposed greater participation in the regulation of CFTR than NBD2.

Class IV mutations affect the conductance of CFTR and have found to be located in MSD1. These are mutations which do not affect the processing or localisation of the mutant protein, but reduce the CFTR current.

There appears to be some correlation between the mechanism by which CF mutations disrupt CFTR function and the severity of pancreatic disease (Welsh and Smith 1993). Sheppard *et al.* (1993) have shown that mutant CFTR containing mutations associated with mild pancreatic disease retains significant apical chloride channel activity when expressed in epithelial cells. This residual chloride channel activity would explain the dominant nature of mild mutations when combined with a severe allele in compound heterozygotes.

Table 1.2 Classes of *CFTR* Mutations That Cause CF

CLASS	DEFECT	EXAMPLES	DOMAIN	FREQUENCY	CLINICAL STATUS
I	NO PROTEIN				
	Nonsense mutations	G542X	NBD1	3.4	PI
	Frameshift mutations	3905 insT	NBD2	2.1	PI
	Splice site mutations	621+G→T	MSD1	1.3	PI
II	PROCESSING	Δi506/7	NBD1	0.5	PI
		ΔF508	NBD1	67.2	PI
		S5491	NBD1	RARE	
		S549R	NBD1	0.3	PI
		A559T	NBD1	RARE	
		N1303K	NBD2	1.8	PI
III	REGULATION	G551D	NBD1	2.4	PI
		G551S	NBD1	RARE	PS
		G1244E	NBD2	RARE	PI
		S1255P	NBD2	RARE	PI
		G1349D	NBD2	RARE	PI
IV	CONDUCTION	R117H	MSD1	0.8	PS
		R334W	MSD1	0.4	PS
		R347P	MSD1	0.5	PS

(Taken from Welsh and Smith 1993).

Frequency of mutations is expressed as percentage of all CF mutations.

NBD refers to nucleotide binding domain, MSD membrane spanning domain, PI pancreatic insufficiency, and PS pancreatic sufficiency.

A1.1.7 The Expression Of CFTR

Cell-specific expression of *CFTR* was studied by Trezise and Buchwald (1991) using labelled antisense RNA probes for *in situ* analysis on rat tissue sections. They found that *CFTR* was specifically expressed in the ductal cells of the pancreas and salivary glands, and that there were decreasing gradients of expression in the intestine along the crypt-villus, proximal-distal axes. *CFTR* expression was also detected at low levels in the mucosa and submucosa of the bronchi and bronchioles of the lung, with the highest expression being detected in the sub-mucosal glands. Expression in the testis was found to be regulated during the cycle of the seminiferous epithelia, and implies that CFTR plays a role in spermatogenesis thereby contributing to CF infertility.

The localisation of CFTR has been studied immunocytochemically using antibodies raised against CFTR. Great difficulty has been experienced in raising antibodies to CFTR and in achieving a uniform pattern of expression. Crawford *et al.* (1991) found abundant CFTR in epithelial cells including those lining the sweat ducts, small pancreatic ducts, and intestinal crypts. The level of CFTR in lung epithelia was again found to be surprisingly low, but was abundant in kidney tubule epithelia. The protein appeared to be restricted to the apical membranes of cells, a feature also described by Cohn *et al.* 1991., Denning *et al.* 1992., and Zeitlin *et al.* 1992., which, along with tissue distribution is consistent with its proposed role as a chloride channel.

Expression of *CFTR* during foetal development has also been studied, using *in situ* techniques to detect mRNA (Tizzano *et al.* 1993., Trezise *et al.* 1993). This demonstrated expression of *CFTR* in all major organs known to be affected in CF, with an expression pattern which broadly matched that of the adult. However, these studies also identified some interesting differences between adult and foetus *CFTR* expression. The level of *CFTR* expression in the foetal lung epithelium was higher than that seen in adults, and led to the suggestion that *CFTR* might fulfil a different role in the aqueous environment of the foetal lung (Tizzano *et al.* 1993). In addition,

there was a conspicuous lack of expression in the foetal sub-mucosal glands, which is the site of highest expression in the adult. Expression of *CFTR* in reproductive tissues varied between males and females. No expression was detected in the first and second trimester of foetal development in females, however significant expression was apparent in uterine, cervix and fallopian tube epithelia in the third trimester. In males, a low level of expression was detected in the epididymis epithelia at all stages. However, no expression of *CFTR* was detected in ovary or testis at any stage of foetal development.

Although *CFTR* has been shown to be expressed in all tissues affected by the disease process of CF, the levels of expression do not appear to correlate with the degree of pathology. However, there is no reason to presume that the major sites of disease pathology would exhibit the highest levels of expression, as each tissue and organ might have differential expression requirements (Porteous and Dorin 1993). Alternatively, the variation in disease pathology might be more a reflection of varying tissue vulnerability to defective chloride transport, or the consequent luminar blockage.

A1.1.8 Cross Species Analysis

Cross species analysis of CFTR sequences has demonstrated a high level of conservation indicating that CFTR has been highly conserved throughout evolution. Murine CFTR has an overall identity to human of 78.3% at the amino acid level, which increases to 89.6% if conservative changes are included (Yorifuji *et al.* 1991., Kelley *et al.* 1992). Analysis of rat CFTR indicates a nucleotide sequence identity of 80.5%, and an amino acid sequence identity of 75.5% (Trezise *et al.* 1992), which is similar to that of 77.4% reported for *Xenopus laevis* at the amino acid level, and 72% for dogfish CFTR (Marshall *et al.* 1991). Analysis and comparison of many diverse species such as mouse, rat, Guinea pig, rabbit, sheep, cow, pig, dogfish, marmoset, and *Xenopus laevis,* has identified that the highest levels of sequence conservation are always seen in the regions encoding the first membrane spanning domain and the first nucleotide binding domain (Marshall *et al.* 1991., Gasparini *et*

al. 1991., Yorifuji *et al.* 1991., Diamond *et al.* 1991., Kelley *et al.* 1992., Trezise *et al.* 1992., and Tucker *et al.* 1992). The region which is consistently least conserved across species is the R domain. This endorses the predictions of CFTR functional assays which suggest that MSD1 and NBD1 play an important role in regulation of CFTR chloride channel activity through specific interactions, whereas regulation of CFTR by the R domain occurs through a 'steric' non-specific mechanism. In addition it has been observed that in dogfish, mouse, and cow, almost all of the missense CF-causing mutations in humans occur at conserved sites. This high level of cross species conservation, and in particular conservation of the sites of human missense mutations, augurs well for studying the disease in an animal model.